PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/00 — A2

(11) International Publication Number: **WO 00/11144**

(43) International Publication Date: 2 March 2000 (02.03.00)

(21) International Application Number: PCT/GB99/02697

(22) International Filing Date: 13 August 1999 (13.08.99)

(30) Priority Data:

9817959.1	19 August 1998 (19.08.98)	GB
9817963.3	19 August 1998 (19.08.98)	GB
9913014.8	5 June 1999 (05.06.99)	GB

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(71) Applicant: ADVANCED TECHNOLOGIES (CAMBRIDGE) LIMITED [GB/GB]; Globe House, 1 Water Street, London WC2R 3LA (GB).

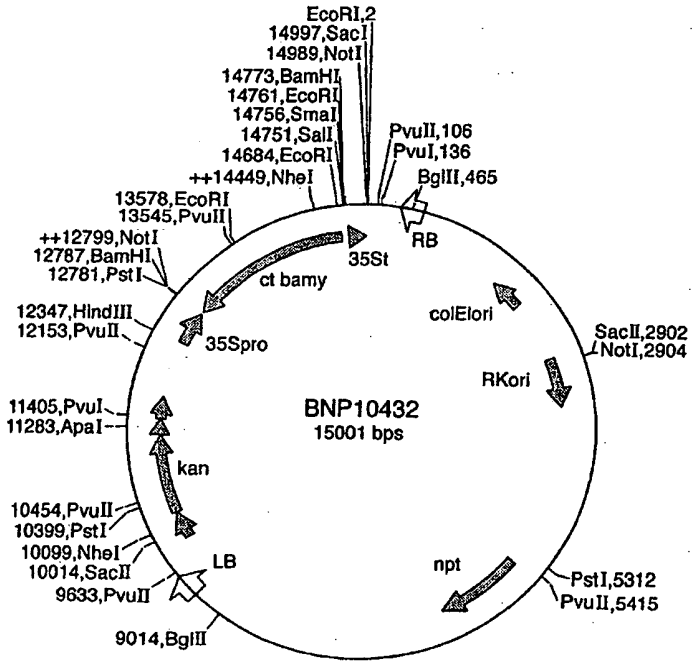
(72) Inventors: KAVANAGH, Thomas, Anthony; Smurfit Institute of Genetics, Lincoln Place, Trinity College, Dublin 2 (IE). LAO, Nga, Thi; Smurfit Institute of Genetics, Lincoln Place, Trinity College, Dublin 2 (IE).

(74) Agents: WALFORD, Margot, Ruth et al.; British American Tobacco (Investments) Limited, Patents Department, R & D Centre, Regents Park Road, Southampton SO15 8TL (GB).

(54) Title: A NOVEL PLASTID–TARGETING NUCLEIC ACID SEQUENCE, A NOVEL β–AMYLASE SEQUENCE, A STIMULUS–RESPONSIVE PROMOTER AND USES THEREOF

(57) Abstract

The invention provides a novel chloroplast targeted novel β–amylase sequence (ct β–amylase), a novel chloroplast targeting nucleic acid sequence and a novel β–amylase sequence. There is also disclosed an inducible promoter which is independently stimulated by light or sugar stimulus. Methods of transforming plants using these sequences are described, as well as transformed plant cells, transformed plants and seed thereof, as well as chimaeric genes containing the sequences. Modification of starch levels in plants can be achieved, as well as the targeting of genes from the starch biosynthetic or degradative pathways, disease or pest resistance or variation of gene expression due to stimulus are described.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

A novel plastid-targeting nucleic acid sequence, a novel β-amylase sequence, a stimulus-responsive promoter and uses thereof

The precise mechanisms by which starch is synthesised and degraded in plants are unknown, despite the isolation and characterisation of a number of enzymes that are presumed to be involved in the process.

Starch is accumulated in the chloroplasts of leaves during the day and is used to supply the needs of the plant for energy and biosynthesis during the night. The mode by which this so-called transient starch is mobilised is not fully understood, but must involve the co-ordinated regulation of synthetic and degradative enzyme activities. In leaf tissues the main degradation pathway is thought to involve phosphorolytic and hydrolytic activities, especially α-glucosidase (E.C. 3.2.1.3)(Nielson and Stitt, 1997).

Starch is also accumulated in the amyloplasts in storage organs such as seeds, fruit and tubers. In this case starch is stored over longer periods of time and mobilisation of the starch is accompanied by degeneration of the storage organ tissues and increases in amylolytic and phosphorolytic activities. However, there is evidence to suggest that turnover of starch is also occurring in the amyloplasts of the storage organ (Sweetlove *et al*, 1996). This again requires the co-

ordinated regulation of the synthetic and degradative enzyme activities.

Chloroplasts and amyloplasts are both derived from proplastids and therefore have many characteristics in common besides being the site of starch synthesis in leaves and storage organs respectively; chloroplasts can be converted to amyloplasts and other types of plastid (Thomson and Whatley, 1980).

Starch is a mixture of two polysaccharides: amylose which is a linear chain of glucosyl units linked by α-1,4-glycosidic bonds; and amylopectin which is made up of many linear chains of α-1,4-polyglucans which are joined together by α-1,6 glycosidic bonds.

Enzymes involved in the synthesis of starch are ADPG pyrophosphorylase (E.C. 2.7.7.21), starch synthase (E.C. 2.4.1.21) and branching enzyme (E.C. 2.4.1.18). ADPG pyrophosphorylase is responsible for supplying the substrate ADPG, this molecule serving as the donor of glucose monomers which are linked together by the concerted action of starch synthases (α-1,4 bonds) and branching enzymes (α-1,6 bonds).

It is thought that the insoluble, crystalline structure of starch grains is formed by the close packing of the extended helical, branched amylopectin molecules, with the linear amylose molecules filling any spaces.

A range of starch-degrading enzyme activities has been

reported including α-amylase (E.C. 3.2.1.1), isoamylase (E.C. 3.2.1.68), β-amylase (E.C. 3.2.1.2), α-glucosidase (E.C. 3.2.1.3), starch phosphorylase (E.C. 2.4.1.1) and disproportionating enzyme (E.C. 2.4.1.25). Many of these enzyme activities exist in multiple forms in plants and some are thought to be involved in the synthesis of starch. All probably take part, to some extent, in the starch mobilisation process, however their exact roles and possible interactions are yet to be determined. The difficulties in attributing roles for the different enzymes is best exemplified by reference to two of the enzyme activities which are thought to be the major contributors to starch breakdown in plants: starch phosphorylase and amylase.

Starch phosphorylase catalyses the reversible release of glucose-1-phosphate from α-1,4-glucans. Two forms of starch phosphorylase are found in plant tissues: Pho1, or the L-type, is located inside plastids and has a high affinity towards maltodextrins; Pho2, or the H-type, is cytosolic and has high affinity to large, highly branched polyglucans such as glycogen. Although the plastidic Pho1 enzyme would be a likely candidate to be involved in the mobilisation of starch, antisense inhibition of the leaf enzyme activity had no effect on the starch accumulation in leaves of transgenic potato plants (Sonnewald *et al.*, 1995). In another study, antisense inhibition of the cytoplasmic Pho2 had an influence on the

sprouting behaviour of transgenic potato tubers, but had no effect on the starch accumulation and degradation (Duwenig *et al.*, 1997).

There are two major groups of amylase both of which hydrolyse α-1,4-glucosidic linkages in amylose and amylopectin: α-amylase acts randomly on non-terminal linkages, whereas β-amylase acts to release maltose units starting from the non-reducing end of the polyglucan chain. The subcellular location of α-amylase in the apoplastic space of plant cells is thought to reflect the fact that the enzyme is normally secreted. However, in a number of plants such as rice (Chen, *et al.*, 1994) and sugar beet (Li, *et al.*, 1992) the enzyme is also located inside chloroplasts and amyloplasts, despite the finding that the signal sequences at the amino-terminus of a number of α-amylase proteins are characteristic for translocation of protein across the ER membrane rather than the plastid membrane (Chen *et al*, 1994). In a study where the promoter and signal sequence of a rice α-amylase gene was fused to the bacterial GUS gene and introduced into rice, tobacco and potato using *Agrobacterium*-mediated transformation (Chan *et al.*, 1994), it was demonstrated that the expressed GUS fusion protein was first transported to the endoplasmic reticulum and then exported into the culture medium of suspension cultures made from transgenic cells. It has been shown in a number of

studies that α-amylase will degrade native starch molecules.

In contrast, *in vitro* studies have shown that β-amylase will not degrade native starch granules without prior digestion of the granule with other enzymes. Mutants of rye (Daussant *et al.*, 1981) and soybean (Hildebrand and Hymowitz, 1981) that lack active β-amylase or contain only traces of activity, respectively, apparently show normal growth and development. In addition, transgenic *Arabidopsis* plants in which the levels Of β-amylase have been greatly reduced, do not show severe growth defects (Mita *et al.*, 1997). Attempts to define the precise physiological role of β-amylases in plants have been hampered by inconclusive data concerning subcellular location. Although one study (Kakefuda *et al.*, 1986) reported the presence of two β-amylases in pea chloroplasts, most studies involving species such as *Vicia faba*, barley, wheat, soybean, sweet potato and pea have concluded that most, if not all, β-amylase activity is extrachloroplastic (Nakamura *et al.*, 1991). This view is supported by the fact that all β-amylase genes cloned to date encode proteins that lack amino-terminal chloroplast transit peptide sequences.

In cereals, three types of β-amylase have been described: an endosperm-specific form that accumulates during caryopsis maturation; a form that is synthesised *de novo* in aleurone cells of rice and maize during germination (Wang *et al.*, 1996;

1997); and a β-amylase which is ubiquitous in vegetative organs. In *Arabidopsis*, the ubiquitous form accounts for approximately 80% of the total starch-degrading activity of rosette leaves. In common with all other β-amylase genes cloned to date, the gene for the ubiquitous *Arabidopsis* β-amylase does not encode a protein with a subcellular targeting signal, thus the enzyme is likely to be located in the cytosol.

The findings from a number of studies that the degradative activities can be removed without an adverse effect on the viability of the plant, plus the subcellular location of starch degrading enzymes outside the plastid, is surprising. The apparent absence of a plastid-localised β-amylase activity is especially surprising in light of the fact that the expected major end-product of β-amylase activity, namely maltose, has been identified as a product of starch degradation in isolated chloroplasts (Peavey *et al.*, 1977). More recently, it has been shown that both glucose and maltose are exported from isolated cauliflower bud amyloplasts during the process of starch mobilisation (Neuhaus *et al.*, 1995).

The ability to manipulate the amount of starch in the plastids of leaves or storage organs would be of high benefit to various industrial processes which utilise plant starches. For example, in an attempt to increase the starch content of potato tubers, it has been shown previously that when *E. coli*

ADPG PPase *glgC*16 is overexpressed in transgenic potato tubers, there is an increase in flux of carbon into starch but there is only a small increase in net accumulation of starch (Sweetlove *et al.*, 1997). Analysis of enzyme activities in the overexpressing lines showed that, apart from the alteration in ADPG PPase, the activity of amylase, specifically β-amylase was also altered. This data suggests that the accumulation of starch in tubers overexpressing glgC16 protein is prevented by the breakdown of the newly synthesised starch, i.e. the starch is being turned over.

In another example, the availability of starch during the malting process is closely correlated with the types and amounts of degradative enzyme activities in the plant, specifically the storage organs. An increase in the degradative capacity of the crop would make the malting of cereal grain or the conversion of starch from tubers, or other storage organs, to alcohol more efficient and productive.

The type of starch present in the storage organ depends on the forms and activities of the ADPG pyrophosphorylase, starch synthase, branching enzyme and the degradative enzymes present. The interactions between the various enzymes will also be important.

There is considerable interest in creating novel starches *in planta* as this will reduce the costs of processing and modification of the starch before use in a variety of

industries such as food, paper, pharmaceuticals, glue, oil and textiles. The following examples show how starch hydrolytic activity can be important in altering the structure of starch *in vivo*.

It has been shown that, in maize kernels, the *sugary1* mutation causes the absence of a debranching enzyme which hydrolises α-1,6-glycosyl linkages of starch (James *et al.*, 1995). The mutation results in the decreased concentration of amylopectin and accumulation of the highly branched glucopolysaccharide, phytoglycogen.

It has been shown that in pea, short oligosaccharide molecules, starting with maltose and adding successive glucose units up to maltoheptose, specifically stimulate the activity of granule bound starch synthase I (GBSSI) (Denyer *et al.*, 1996) which is generally accepted to be the major enzyme responsible for the synthesis of amylose (e.g. van der Leij *et al.*, 1991; Hylton *et al.*, 1995; Ainsworth *et al.*, 1993). The manipulation of GBSSI activity by controlling the supply of malto-oligosaccharides is the subject of a recent patent (WO 97/16554) and suggests that an increase in the concentration of malto-oligosaccharides, and thus an increase in the ratio of amylose to amylopectin in the starch, can be brought about by the introduction of degradative enzymes namely α-amylase, β-amylase, disproportionating enzyme, debranching enzyme and starch phosphorylase. Patent WO 97/16554 also states that genes

for plastidial isoforms of these enzymes have been cloned. However, as discussed above, no β-amylase genes isolated to date encode a β-amylase enzyme with a protein targeting sequence and, in addition, there is doubt that α-amylases are originally targeted to plastids (Chen *et al.*, 1994; Chan *et al.*, 1994). Later in WO 97/16554, reference is made to the engineering of a suitable β-amylase cDNA sequence to add a plastid targeting sequence.

In addition to the industrial uses for starch in the storage organs, the amount of starch in the leaf has significant importance for the agronomy of a crop. Starch is synthesised in the leaf during daylight from the carbon fixed during photosynthesis. The starch is stored in the chloroplast and is broken down at night to become a source of energy and intermediates for metabolism in the plant. By which mechanisms the source-sink relationship is controlled are unknown at present, however, it is clear that manipulation of the amount and availability of the starch in leaf plastids will have a profound influence on plant productivity (biomass and yield).

The amount of starch in the leaf will also be important for those crops where the leaf is the major plant commodity, for example tobacco. It is known that starch content has an influence on the eventual flavour of tobacco when smoked. Provision of a means to manipulate the level of starch in

tobacco leaves could be of interest to the tobacco industry.

We describe here, for the first time, the isolation of a cDNA encoding a novel β-amylase enzyme which is targeted to plastids (henceforth known as chloroplast targeted (ct) β-amylase), by a novel targeting sequence. The isolation of this entire coding sequence is surprising, as it has generally been thought that β-amylase would only take part in the hydrolysis of starch once smaller polyglucan fragments had been released, either by translocation or through breakdown of the membrane, from the plastid into the cytoplasm. Location of the enzyme in plastids opens up the unforeseen possibility that ct β-amylase is involved in the degradation of transient starch located in chloroplasts and storage starch located in amyloplasts.

The similarity of characteristics between chloroplasts and amyloplasts (Thomson and Whatley, 1980) is of relevance to the current invention, as it has been shown that the transit peptides from chloroplast-targeted polypeptides can import heterologous polypeptides into amyloplasts and *vice versa*. For example, the transit peptide from the maize granule bound starch synthase enzyme when fused to the *E. coli* β-glucuronidase (GUS) protein will import the GUS protein not only into amyloplasts but also into chloroplasts (Klosgen and Weil, 1991).

In addition, we show that expression of the *ct-Bmy* gene in

Arabidopsis and the expression of *ct-Bmy* promoter:GUS fusions in transgenic tobacco can be regulated independently by both light and sucrose. This is surprising in view of the tightly coupled light and sugar induction responses of *ATß-Amy* of *Arabidopsis* (Mita *et al*, 1995).

The present invention provides a nucleic acid sequence known herein as SEQ. ID. No. 1 and being from 1-294 nucleotides and having therewithin a sequence capable of targeting a further coding sequence to a plant plastid, or sequences being at least 65% or more homologous with the disclosed sequence SEQ. ID. No. 1 and having the same targeting ability.

Preferably the nucleic acid sequence encodes about 94 and more preferably about 85 amino acid residues.

The present invention also provides a nucleic acid sequence known herein as SEQ. ID. No. 2 and being from 1-1642 nucleotides and having therewithin a sequence capable of encoding β-amylase, or sequences being at least 65% or more homologous with the disclosed sequence within SEQ. ID. No. 2 and having the same encoding ability.

The present invention also provides a nucleic acid sequence known herein as SEQ. ID. No. 3 and being from 1-1953 nucleotides and having therewithin a sequence capable of encoding chloroplast targeted β-amylase, or sequences being at least 65% or more homologous with the disclosed sequence within SEQ. ID. No. 3 and having the same encoding ability.

Homologous sequences also include those sequences which hybridise to SEQ. ID. No. 1, SEQ. ID. No. 2 or SEQ. ID. No. 3 under medium stringency conditions (washing at 2x SSC at 65°C).

Preferably the nucleic acid sequence is an mRNA or cDNA sequence, although it may be genomic DNA.

The present invention also provides a method of increasing or decreasing in a plant the activity of an enzyme in the pathway of starch biosynthesis or degradation, the method comprising the steps of stably incorporating into a plant genome a chimaeric gene comprising a nucleic acid sequence encoding a plastid targeting sequence and a coding sequence for an enzyme in the starch biosynthetic or degradative pathway, and regenerating a plant having an altered genome.

The present invention also provides a method of targeting proteins or enzymes to a plant plastid, the method comprising the steps of stably incorporating into a plant genome a chimaeric gene comprising a nucleic acid sequence encoding a plastid targeting sequence and a coding sequence for a protein or an enzyme, and regenerating a plant having an altered genome, the protein or enzyme being one or more in the pathway of the following group: lipid synthesis, photosynthesis, amino acid metabolism, nitrogen fixation, carbon fixation or synthesis of carbohydrate polymers; or being able to confer a characteristic to the plant, the characteristic being selected from one or more of the following group: herbicide resistance

and pest resistance, for example, including fungal, bacterial or viral resistance.

The present invention also provides plants having therein a chimaeric gene comprising a promoter, a nucleic acid coding sequence encoding the plastid targeting sequence, the sequence being capable of targeting a coding sequence of an enzyme in the starch biosynthetic or degradative pathway to a plant plastid, and a terminator.

The present invention further provides a nucleic acid sequence capable of directing expression of a product encoded by a coding sequence which is operably linked thereto, said nucleic acid sequence being known herein as SEQ.ID. No. 8, or being at least 65% homologous therewith and having substantially the same function thereas, and said nucleic acid sequence being responsive to stimulus, the level of expression of said product being variable in response to the stimulus applied to said nucleic acid sequence.

The present invention further provides a method of varying the level of expression of a product encoded by a coding sequence operably linked to a nucleic acid sequence capable of directing expression of said product in a plant, said method comprising the steps of stably incorporating into a plant genome a chimaeric gene comprising a nucleic acid sequence capable of directing expression of a product encoded by a coding sequence that is operably linked thereto, said

nucleic acid sequence having substantially the sequence of SEQ.ID. No. 8 or being at least 65% homologous therewith and having substantially the same function thereas, and being responsive to stimulus.

Preferably the stimulus is the presence or absence of light and/or varying levels of sugar. Alternatively the stimulus is a stimulus which is developmentally controlled.

Advantageously the sugar is one or more of sucrose or glucose.

Preferably the sugar is sucrose.

Advantageously the inducible promoter, or nucleic acid sequence capable of directing expression of said product in a plant, is operable under conditions when there is no light but sugar is present, or when there is no sugar but light is present. The tissue of a plant where no light but sugar is present may suitably be underground organs or sink organs. Underground organs may be, for example, tubers, rhizomes or roots, whereas other sink organs may be young leaves or seeds.

The tissue of a plant where no sugar but light is present may be older leaves (where no sugar is transported), flower parts or germinating seeds.

Constructs and chimaeric genes having the DNA structural features described above are also aspects of the invention.

Plant cells containing a chimaeric gene comprising a nucleic acid sequence encoding a plastid targeting sequence

hereinabove described and a nucleic acid coding sequence of an enzyme in the starch biosynthetic or degradative pathway, or a chimaeric gene comprising a nucleic acid sequence capable of directing expression of a further coding sequence, or a chimaeric gene comprising a nucleic acid sequence hereinabove described that is responsive to stimulus and a coding sequence, the level of expression of said coding sequence being variable in response to the stimulus applied to said nucleic acid sequence are also an aspect of this invention, as is the seed of the transformed plant containing one or more chimaeric genes according to the invention.

Advantageously the plastid targeting sequence is the sequence SEQ. ID. No. 1.

In a first aspect of the invention the above method may be used to alter the metabolism of a leaf such that starch is accumulated therein or mobilised therefrom, this process altering the source-sink relationships within the plant as a whole. Such may be achieved by providing the targeting sequence and a nucleic acid coding sequence of an enzyme in the starch biosynthesis or degradative pathway under the direction of a suitable promoter. Suitable promoter selection would result in plants with increased or decreased levels of starch in the leaves which might be useful, for example, in the tobacco industry; or alternatively would result in changes in yield of starch in various other plant tissues such as tubers,

fruit and roots following modification of the source-sink relationships of the plant.

In this embodiment of the invention a suitable promoter would direct expression of the plastid targeting sequence and the coding sequence of an enzyme in the starch biosynthetic or degradative pathway throughout the whole plant, so called constitutive expression, or specifically to the leaves. These changes will have a profound effect such that the starch content and/or the yield of the organs of the plant would be significantly altered.

A preferred promoter capable of directing expression throughout all plant tissues is the promoter taken from cauliflower mosaic virus 35S gene. For leaf expression, preferred promoters can be taken from the gene for the small subunit of ribulose bisphosphate carboxylase or the pea plastocyanin gene. One skilled in the art will recognise other suitable promoters both for constitutive expression and specific leaf expression such as the nopaline synthase promoter and the chlorophyll a/b binding protein promoter respectively.

The coding sequence, or parts thereof, for the enzyme in the starch biosynthetic or degradative pathway may be arranged in the normal reading frame direction, i.e. sense, or in the reverse reading frame direction, i.e. antisense. Up or down regulation of the activity of the enzyme in a plant using sense, antisense or cosuppression technology (the latter as

described by DNAP in their European Patents Nos. 0465572 and 0647715) may be used to achieve alteration in the starch of the plant.

In a second aspect of the invention the inventive method may also be used to alter the metabolism of starch in storage organs such that starch content is increased and/or the starch is provided in a suitable form as required for the purposes of particular industrial processes. Such processes including paper making; manufacture of pharmaceuticals, textiles, dyes and building products; provision of baking, dairy and snack food products; making canned, dried or instant foods; malting of grain and production of syrups and alcohol.

In the first or second aspect of the method the enzyme selected for use in the chimaeric gene of the methods may be one from the starch degradative pathway, i.e. a starch degrading enzyme. Advantageously, the chimaeric gene comprises a chloroplast targeted β-amylase (hereinafter known as ct β-amylase), and more preferably comprises ct β-amylase derived from *Arabidopsis thaliana,* (hereinafter known as At ct β-amylase), see SEQ. ID. No. 3. Sequences homologous to At ct β-amylase which may be derivable from other plant sources such as potato, tobacco, wheat, maize and barley may also be used. Standard methods of cloning by hybridisation or polymerase chain reaction (PCR) techniques may be used to isolate

sequences from such organisms: for example molecular cloning techniques such as those described by Sambrook *et al.* (1989) and the PCR techniques described by Innes *et al.* (1990). Other starch degrading enzymes, the coding sequence of one or more of which would be suitable for use with the plastid targeting sequence, include α-amylase, disproportionating enzyme, debranching enzyme, starch phosphorylase, α-glucosidase and non-plastidic β-amylase.

In the second aspect of the inventive method preferred promoters which would direct expression to the storage organs of plants could be selected, for example, from the genes from the following list: the gene for high molecular weight glutenin of wheat endosperm; the gene for α,β-gliadin of wheat endosperm; the hordein gene of barley endosperm; or the gene for patatin from potato tubers. Other suitable promoters are known to those skilled in the art.

In either aspect of the invention, the alteration of tissue metabolism or alteration of starch type or characteristics may be made stimulus responsive, i.e. inducible, by virtue of use of the inducible promoter described herein (SEQ. ID. No. 8). For example, the light inducibility aspect of the inducible promoter could be used to manipulate seed set by inducing a gene such as Barnase (as exemplified in Patent WO 98/10081) to affect pollen development, or to affect

non-light responsive genes in otherwise light-dependant processes such as fruit ripening or seed germination. The light inducible promoter could also be used to turn on genes which affect secondary metabolite production in leaves, for example alkaloid production. Light inducible promoters may also be used to manipulate starch biosynthetic enzyme genes in leaves or other photosynthetic tissue, or for example in turning on genes after removal of tubers, for example, from storage in darkness. The sugar-inducibility aspect of the inducible promoter could be used to regulate genes in, for example, developing tuber or other non-photosynthetic tissue such as genes for pest resistance and/or genes which might affect the quality of the post-harvest crop. For potatoes, resistance genes to blight, blackleg and dry rot would be particularly of benefit and could be most advantageously cloned into recombinant genes with sugar inducible promoters. Alternatively, the sugar inducibility aspect of the inducible promoter could be used to drive the expression of genes for selectable markers in the tissue culture process.

One skilled in the art can readily delineate the sugar inducible responsive element from SEQ. ID. No. 8 and/or the light inducible responsive element by using well known techniques, such as deletion studies. Pwee and Gray (1993) describe such a deletion study within the pea plastocyanin gene using a marker gene in order to determine operative regions

thereof.

Methods described herein or in, for example, laboratory manuals by Sambrook et al (1989) and Gelvin and Stanton (1995) for cloning gene sequences and inserting them into appropriate carriers (vectors or plasmids etc.) are techniques well known to the skilled man for putting such concepts into effect. The chimaeric gene or genes as described above may be introduced on their own, or be accompanied by one or more other chimaeric genes, such as one or more of the other genes described above. In the case of the above described embodiments utilising a first chimaeric gene encoding an enzyme of the starch degradative pathway, the second chimaeric gene may, for example, comprise a nucleic acid sequence encoding an enzyme from the starch biosynthetic pathway also under the direction of a suitable promoter and a suitable terminator. The promoter and/or terminator of the second chimaeric gene may be the same as or different from the promoter and/or terminator of the first chimaeric gene. Suitable sequences encoding enzymes from the starch biosynthetic pathway are the nucleic acid sequences for sucrose synthase, ADPG pyrophosphorylase, starch synthase, and may also include branching enzyme, α-amylase, isoamylase, non-plastidic β-amylase, α-glucosidase, starch phosphorylase and disproportionating enzyme.

Methods for the introduction of more than one chimaeric gene into a plant have been described and comprise the

construction of a binary vector with the chimaeric genes joined together in one nucleic acid molecule; cotransformation using two or more different *Agrobacterium* cells, for example, with different binary vectors containing different chimaeric genes therein; or the transformation of a plant which already has a chimaeric gene with a second, different chimaeric gene, i.e. retransformation. In the latter case, the method of selection of transgenic plants after the introduction of the second chimaeric gene must be different from the selection method used for the introduction of the first chimaeric gene. Suitable selectable markers would include those for hygromycin, kanamycin, sulphonamide and Basta resistance. Biological methods such as crossing two plants, each plant containing a single chimaeric gene can also be used.

Use of two chimaeric gene constructs could be made in order to alter the starch content of an already transformed plant which shows a significant increase in a first enzyme activity and a consequent change in the synthesis of starch.

Thus, the present invention further provides a method of altering in a transgenic plant, which plant already shows an increase or decrease in an enzyme activity as a result of genetic transformation, a further enzyme in order to up or down regulate said further enzyme and thereby increase or decrease the amount of starch produced by the retransformed plant.

Advantageously the first transformed plant is a plant

having an increased enzyme activity in the starch biosynthetic pathway. An example of an attempt to increase the starch content of a plant is a transgenic potato transformed with the gene for ADPG-PPase, for example *glgC16 (see for example, WO 91/19806)*. The amount of starch increase in such a plant has been relatively small. This first transformed plant is advantageously retransformed with a chimaeric gene for a starch degrading enzyme, suitably comprising, for example, At ct β-amylase. The glgC16 protein is expressed in the first transformed tubers and results in increased ADPG-PPase activity and an increase in flux of carbon to starch. Advantageously, the expression of the chimaeric At ct β-amylase gene, or parts thereof, in the retransformed tubers results in down regulation of the ct β-amylase activity, i.e. cosuppression or antisense technology, thus providing for an increase in starch accumulation.

Preferably the expression of the second enzyme is directed to tubers. A suitable promoter to direct the expression of the At ct β-amylase chimaeric gene in tubers is the promoter from the gene for patatin.

The first transformed potato plant expressing *glgC16* is kanamycin resistant, therefore the binary vector construct for the At ct β-amylase chimaeric gene carries a different resistance gene, suitably a gene for sulphonamide resistance,

for example. Increased starch production in the potato tuber would be of benefit, for example, to the potato crisp manufacturer as a 1% increase in potato dry matter would result in a 4% increase in product.

Potato crisp manufacture also serves to illustrate another benefit of the invention. When potato tubers are stored at temperatures below 8°C, reducing sugars, glucose and fructose from the breakdown of starch accumulate. When the potatoes are fried for crisps the reducing sugars react with amino acid in the Maillard reaction to give rise to brown colouration and off-tastes in the product. Introduction into potato plants of a chimaeric gene which would stop the breakdown of starch and thus the accumulation of reducing sugars would be of benefit to the snack food industry. Preferably the chimaeric gene would comprise the coding sequence, or a part of the sequence, for ct β-amylase in a cosuppression or antisense construct, driven by a suitable promoter and terminator. A suitable promoter would be taken from the gene for patatin in potato tubers. Advantageously any of the other starch degrading enzymes mentioned above could also be used instead of the ct β-amylase.

The inducible promoter of SEQ. ID. No. 8 could also be used in the construct if co-ordinated expression in the developing leaf and in the developing tuber were required, as the patatin promoter is also sucrose inducible (Rocha-Sosa *et al* (1989). Similarly, the sequence for the chloroplast

targeting polypeptide of SEQ. ID. No. 1 could also be used with any other gene which lacked its own targeting sequence and which was required to be directed to plastids.

The above examples serve to illustrate the possible benefits of using the present invention. One skilled in the art will recognise that the combination of genes and the plants to which the invention could be applied is considerable.

Gene combinations preferably will include ct β-amylase with one or more of the genes for sucrose synthase, ADPG pyrophosphorylase, starch synthase, branching enzyme, α-amylase, isoamylase, non-plastidic β-amylase, α-glucosidase, starch phosphorylase and disproportionating enzyme, the sequences of which are known to the skilled man. Alternatively, the targeting sequence from ct ß-amylase may be used with one or more of the above genes.

The list of plants which could be transformed preferably include potato, wheat, maize, barley, tomato, rice, pea, soybean, peanut, cassava, yam, banana and tobacco.

The invention will now be described, by way of example, with reference to an embodiment for isolation of the cDNA for ct β-amylase from *Arabidopsis thaliana* and for incorporating the cDNA into tobacco and potato plants. Examples are also given on the stimulus responsive promoter and its activity in transgenic plants.

In order that the invention may be readily carried into effect reference will now be made, by way of example, to the following diagrammatic drawings in which;

Figure 1 shows the results of radiolabelled *in vitro* import translation products sampled on SDS-PAGE gel followed by fluorography. Legend: Molecular weight markers (lane M); translation products (lane Tr); chloroplasts reisolated and thermolysin-treated after import incubation (lane C); stromal fraction (lane S); washed thylakoids (lane T); thermolysin-treated thylakoids (lane tT); inner envelope fraction (lane I); outer envelope fraction (lane O). Putative precursor (P), intermediate (I) and mature (M) forms of β-amylase respectively. KiloDaltons (K);

Figure 2 shows the effect of light and the effect of light and sugars on the expression of ct β-amylase transcript in Arabidopsis thaliana seedlings. Figure 2a shows Northern blot analysis of total RNAs of 5-week old *Arabidopsis* plants grown in soil and exposed to 2 days continuous light (L), 2 days continuous darkness (D), 2 followed by 3 days of continuous light (LL) or 2 days of darkness followed by 3 days of continuous light (DL). Figure 2b shows Northern blot analysis of total RNAs of 5-week old Arabidopsis plants, grown *in vitro*, which were transferred either into water and exposed to 3 days continuous light (WL); or into 5% sucrose and exposed to 3 days of darkness (SD) or 3 days of continuous light (SL);

or into 5% glucose and exposed to 3 days of darkness (GD) or 3 days of continuous light (GL). Northern blots were hybridised with a radiolabelled *ct-Bmy* cDNA insert and autoradiographed (upper panels). The corresponding ethidium bromide-stained formaldehyde-agarose gels are shown in the bottom panels;

Figure 3 shows the diagrammatic representation of the T-DNA of the chimaeric ct β-amylase promoter-GUS genes constructed in Example 3 below, in which NosP represents the nopaline synthase promoter; NosT represents the nopaline synthase terminator; BR is the right border inverted repeat and BL is the left border inverted repeat of the T-DNA of pBI101; NPTII represents the neomycin phosphotransferase II coding sequence; GUS represents the β-glucuronidase coding sequence. ct β-amylase promoter fragments are represented by hatched rectangles; the PCR amplified Xho I-Bam HI bridging fragments are represented by black rectangles;

Figure 4 shows the effect of light and sucrose on the GUS activity expressed from a *ct Bmy* promoter-GUS chimaeric gene in tobacco seedlings;

Figure 5 shows the plasmid map of donator vector pDV35S (SK)V;

Figure 6 shows the plasmid map of donator vector pDV02000;

Figure 7 shows the plasmid map of binary plasmid pBNP10431 where 35Sp represents the CaMV 35S promoter, ct bamy represents the full length ct ß-amylase cDNA, 35St represents the CaMV 35S

terminator, RB represents the right border of the binary vector pBinPlus, colE1ori represents the colE1 origin bacterial replication, RKori represents the oriV origin of replication of the RK2 plasmid, nptIII represents the neomycin phosphotransferase gene for bacterial resistance to kanamycin, LB represents the left border sequence of the binary vector, and kan represents the plant neomycin phosphotransferase recombinant gene required for plant resistance to kanamycin;

Figure 8 shows the plasmid map of binary plasmid pBNP10432 where abbreviations are as for Figure 7;

Figure 9 shows the plasmid map of binary plasmid pBNP02431 where abbreviations are as for Figure 7 except that patp represents the patatin class I promoter from vector pDV02000 in Figure 6 and nost represents the nopaline synthase terminator; and

Figure 10 shows the plasmid map of binary plasmid pBNP02432 where abbreviations are as for Figure 9.

In the sequence listing:

SEQ. ID. No. 1 is the nucleic acid capable of targeting a coding sequence to a plant plastid, particularly a chloroplast;

SEQ. ID. No. 2 is the nucleic acid which encodes β-amylase;

SEQ. ID. No. 3 is the complete sequence of chloroplast targeted (ct) β-amylase;

SEQ. ID. Nos. 4 and 5 are primers used in the

amplification process of Example 3;

SEQ. ID. Nos. 6 and 7 are primers used in the amplification process of Example 4; and

SEQ.ID. NO. 8 is the nucleic acid which is stimulus responsive, particularly to light and/or sugar.

Example 1

Isolation and characterisation of *Arabidopsis thaliana* chloroplast targeted β-amylase

Sequencing of cDNA insert in pBmy81

A BLASTN database search of the nucleotide sequence of a 37kb *Arabidopsis* chromosome IV DNA fragment in cosmid G16599 (Bevan *et al.*, 1998) revealed the presence of a gene sharing significant homology with the extrachloroplastic β-amylase of *Arabidopsis*, barley maize rice soybean and rice. The search also identified several 3' terminal EST sequences, one of which, EST 81E10T7 (Newman *et al.*, 1995), hereafter referred to as pBmy81, was identical over approximately 300 nucleotides. Clone EST 81E10T7 was supplied by the Arabidopsis Biological Resource Center (ABRC) DNA Stock Center (Ohio University, USA). A nested set of *Bal*31 deletion subclones, spanning the cDNA insert in pBmy81, were used as DNA templates in double stranded PCR cycle sequencing reactions using fluorescent dye-labelled universal primers. Sequencing reactions were analysed on an

Applied Biosystems Model 373A automated sequencer. The nucleotide sequence of the cDNA insert in pBmy81 is shown in SEQ. ID No. 3. The construct pBmy81 was deposited by Advanced Technologies (Cambridge) Limited of 210 Cambridge Science Park, Cambridge CB4 4WA under the Budapest Treaty on the International Recognition of the Deposit of Micro-Organisms for the purposes of Patent Procedure at the National Collection of Industrial and Marine Bacteria (NCIMB), 23 St. Machar Street, Aberdeen Scotland on 4 August 1998 under Accession No. NCIMB 40964.

Identification of a putative chloroplast targeting signal

The pBmy81 cDNA insert comprises 36 untranslated nucleotides at the 5' end, an open reading frame (ORF) that encodes a protein of 548 amino acids and a 3' untranslated region (UTR) of 232bp. The protein encoded by the pBmy81 cDNA insert has a predicted molecular weight of 61 kDa and shares high amino acid similarity with plant extrachloroplastic β-amylases from maize, rice, barley, soybean and sweet potato. However, the protein encoded by pBmy81 differs from all other β-amylases reported so far in that it contains a unique N-terminal extension possessing the characteristics of a chloroplast targeting signal i.e. a high content of serine (16%), threonine (10%) and positively charged amino acid

residues (15%) (Baier and Dietz, 1997). Three domains which are distinguishing features of chloroplast targeting signals (Schatz and Dobberstein, 1996) were identified in the signal sequence: an uncharged amino-terminal domain; a central domain rich in hydroxylated amino acids; and a carboxy-terminal domain with the potential to form an amphiphilic β-strand.

cDNA insert in pBmy81 encodes a chloroplast targeted β-amylase

Intact chloroplasts were isolated from 50-60 g of pea shoots (*Pisum sativum* L. var Feltham First) using Percoll step-gradients. Plant material was grown and chloroplasts isolated according to the method of Mould and Gray (1997a).

The pBmy81 plasmid was linearised by restriction digestion with *Not*I and was transcribed *in vitro* using T7 RNA polymerase. Radiolabelled precursor protein was synthesised in a wheat germ translation system, including ^{35}S-methionine and ^{35}S-cysteine from transcripts of the pBmy81 cDNA essentially as described by Mould and Gray (1997b).

Import of radiolabelled *in vitro* translation products was performed as described by Mould and Gray (1997b). After the import incubation, intact chloroplasts were treated with thermolysin (0.2 mg/ml final concentration in import buffer) for 30 min on ice and then the protease reaction was stopped by the addition of EDTA to 50mM in import buffer. Chloroplasts were re-isolated through a cushion of 40% Percoll in import

buffer and then washed in import buffer (Mould and Gray, 1997b). An aliquot (1/10) of the thermolysin-treated chloroplast sample was taken for analysis and the remainder was fractionated essentially as described by Schnell and Blobel (1993). Samples of thermolysin-treated chloroplasts, stromal fraction, thylakoids and thermolysin-trated thylakoids were quantified by SDS-PAGE followed by coomasie blue staining and scanning densitometry of stained protein bands (subunits of ribulose bisphosphate carboxylase and light harvesting complex proteins were used as standards). Equivalent amounts of these fractions (approximately equal to 2% of the chloroplasts recovered from the Percoll gradient), and 505 of the inner and outer envelope fractions recovered, were analysed by electrophoresis on a 10% polyacrylamide gel in the presence of SDS, followed by fluorography. Results (Figure 1) show that the major translation product (lane Tr) was approximately 58 kDa. When isolated, intact pea chloroplasts were incubated with the radiolabelled protein in the presence of ATP, polypeptides of approximately 50 kDa and 48 kDa were generated (lane C). The resistance of these polypeptides to degradation by exogenously added thermolysin, indicates that they are products of radiolabelled protein import. Fractionation of the intact thermolysin-treated chloroplasts into stroma, washed thylakoids, thermolysin-treated thylakoids, inner envelopes and outer envelopes, demonstrated that the two radiolabelled

polypeptides were located in the stromal fraction.

EXAMPLE 2

Sucrose and light induction of *Arabidopsis thaliana* ct β-amylase gene.

To demonstrate the induction of ct b-amylase in light Arabidopsis thaliana ecotype Landsberg plants were grown in the green house under an 18 hour light, 6 hour dark regime at 18°C. After 5 weeks, two trays of seedlings were transferred to complete darkness and two trays of seedlings were grown in continuous light. After two days, one tray of dark-adapted seedlings and a tray of light grown seedlings were used for isolation of total RNAs, and the second tray of each were exposed to a further 3 days of continuous light.

For combined sucrose-light-dark treatments, seeds of Landsberg ecotype were surface sterilised, placed on MS agar medium containing 1% sucrose and grown in a culture room with an 18 hour light, 6 hour dark regime. Five week old seedlings were transferred onto sterilised distilled water or a 5% solution of sucrose or glucose in water. The seedlings were maintained either in continuous light or darkness for three days. Total RNAs were prepared from seedlings of each test and were analysed by northern blot analysis as described by Eggermont *et al.* (1996). Northerns were probed with the gel-purified cDNA insert in pBmy81 following random labelling with

^{32}P-dCTP as described by Feinberg and Vogelstein (1983).

The results shown in Figure 2A indicate that the ct β-amylase genetranscript is inducible due to light.

The results shown in Figure 2B indicate that the ct β-amylase transcript is induced in the dark with 5% sucrose and to a lesser extent with 5% glucose. This induction is enhanced further in the light in the presence of the sugars. These results show that the effect of light and sugars are independent of each other.

EXAMPLE 3

Construction of ct β-amylase promoter-GUS fusions

Promoter fragments were isolated from the ct β-amylase gene located in cosmid G16599 (Bevan *et al.*, 1998) by restriction enzyme digestion. Convenient restriction sites in the promoter were *Hind* III at nucleotide position -1662 bp (starting at 19179bp on the minus strand of SEQ. ID. No. 8), *Sal* I at -1127 bp and *Pst* I at -371 bp and an *Xho* I site located at position +21 bp downstream of the ct β-amylase initiating methionine were used to isolate three different lengths of promoter plus transit peptide sequence (the A of the translation initiation methionine ATG is numbered +1).

A 294 bp (SEQ. ID. No. 1) fragment of the ct β-amylase gene located in cosmid G16599 (Bevan *et al.*, 1998) was amplified using the oligonucleotide primers:

SEQ. ID. No. 4

P1: (5' - AAT TCC TCG AGT TCT CTT ATC - 3') and

SEQ. ID. No. 5

P2: (5' - cgg gAT CCC TGA CAT TGT TAC - 3').

In primer P1, the underlined bases refers to the *Xho* I site located at position +21 bp; in primer P2 the bases in lower case refer to the nucleotides added in order to create a *Bam* HI site.

Chimaeric ct β-amylase promoter-GUS genes were created by triple ligation of the promoter fragment; the PCR bridging fragment digested with *Xho* I and *Bam* HI; and the GUS vector pBI101 (Jefferson *et al.*, 1987) digested with *Hind* III-*Bam* HI, *Sal* I-*Bam* HI or *Pst* I-*Bam* HI (Figure 3). Constructs were termed HβGUS, SβGUS and PβGUS respectively.

The chimaeric gene constructs were transferred to *Agrobacterium tumefaciens* LBA4404 by triparental mating (Bevan, 1984) and introduced into *Nicotiana tabacum* var Samsun by the leaf disk transformation method (Horsch *et al.*, 1985).

EXAMPLE 3A

Sucrose and light induction of chimaeric *Arabidopsis thaliana* ct β-amylase promoter-GUS gene in tobacco seedlings.

Plants containing the HβGUS and PβGUS constructs expressed high levels of GUS activity and F1 seedling progeny of the lines were used to investigate light and sucrose inducible expression of the chimaeric genes. F1 tobacco seeds were surface sterilised, placed on MS agar medium containing 1% sucrose and grown in a culture room with an 18 hour light, 6 hour dark regime. Two to three week old seedlings were transferred onto a 5% sucrose solution or onto distilled water and maintained either in continuous light or darkness for three days. Total protein extracts from pools of 10 to 14 seedlings were analysed for GUS activity using the fluorogenic substrate 4-methylumbelliferyl-glucuronide (4-MUG) as described by Jefferson *et al.* (1987). With both constructs, the level of GUS activity in seedlings exposed to continuous light in the absence of sucrose was similar to the levels of GUS activity in seedlings exposed to sucrose in the absence of light (Figure 4). However, exposure of seedlings to both continuous light and sucrose increased levels of GUS activity by approximately two to three fold. These results are broadly in agreement with the results from the experiments with the ct b-amylase gene itself which showed that light inducibility and sucrose inducibility are independent processes.

Histochemical staining for GUS showed that activity was detected in the cotyledons of two week old seedlings and little or no activity in the first true leaves or in the stems and roots. In four week old seedlings, additional GUS activity was shown throughout the first true leaves and also in the stems. GUS staining was particularly associated with chloroplast-rich parenchyma (chlorenchyma) cells located between the xylem rays and between xylem and the bundles of phloem that constitute the internal phloem in stems.

EXAMPLE 4

Construction of ct β-amylase plasmids for use in transformation of tobacco and potato leaves

Site-directed mutagenesis was used to convert the *Kpn* I site located at position 2302 bp of the pBmy81 plasmid to a *Bam* HI site. Oligonucleotide primers

SEQ. ID. No. 6

P3: (5'- GCT GGT ACG CCT GCA GGA TCC GGT CCG GAA TTC CC - 3')

and

SEQ. ID. No. 7

P4: (5'- GGG AAT TCC GGA CCG GAT CCT GCA GGC GTA CCA GC - 3')

were designed and used with the Quick Change site-directed

mutagenesis kit (Promega). Protocol was as outlined by the manufacturer.

The full length ct β-amylase coding sequence was excised from the mutated pBmy81 plasmid by cleavage with *Bam* HI and then purified with GeneClean (BIO 101). The *Bam* HI fragment was ligated into the *Bam* HI site of the donator vectors pDV35S(SK)V (see Figure 5) and pDV02000 (see Figure 6). pDV35S(SK)V consists of pBluescript (Stratagene) carrying a 35S CaMV promoter-35S terminator, similar constructs are known in the art (e.g. Odell *et al.*, 1985). pDV02000 consists of pBluescript with a 1.4 kbp patatin promoter-nopaline synthase terminator. One skilled in the art could make similar constructs from known sequences (e.g. Liu *et al.*, 1990). Plasmids with the coding sequence in both the sense and antisense orientation relative to the promoters were isolated, and the ct β-amylase chimaeric genes subcloned from the donator vectors into the binary vector pBinPlus (van Engelen *et al.*, 1995). The plasmid maps are shown in Figures 7-10.

EXAMPLE 5

Transformation or Retransformation of plants

Potato plants were transformed using the method of leaf disk cocultivation as essentially described by Horsch (1985). The binary vectors as described above were transferred to

Agrobacterium tumefaciens LBA4404 using the method of electroporation, and cultures of said *Agrobacteria* used in transformation so that regenerated plants carry the chimaeric genes as described in Example 4.

The patatin promoter-ct β-amylase-nopaline synthase terminator chimaeric gene binary plasmid, can be used to transform a potato plant already carrying the chimaeric gene for *E. coli* ADPG-Ppase *glgC*16 by the methods of leaf disk cocultivation.

EXAMPLE 6

Construction of plasmids with the targeting peptide of AT ct β-amylase

The plastid targeting sequence of AT ct β-amylase is contained within a 294 bp fragment equivalent to SEQ. ID. No. 1. PCR amplification or restriction enzyme digestion can be used to isolate fragments of DNA from the plasmids described in Example 3, i.e. fragments will consist of the 35S CaMV promoter plus plastid targeting sequence or the patatin promoter plus the plastid targeting sequence. Chimaeric genes can be constructed by ligating coding sequences for proteins or enzymes as translational fusions with the transit peptide sequence. Translated proteins would be transported to the plastids to provide novel activities or to affect metabolic pathways.

References

Ainsworth, C., Clark, J. and Balsdon, J. (1993). Plant. Mol. Biol., **22**, 67-82.

Baier, M. and Dietz, K.J., (1997) Plant J. 12, 179-190

Bevan, M.W. (1984) Nucl. Acids. Res., 12, 8711-8721

Bevan, M.W. *et al.* (1998). Nature, **391**, 485-488.

Chan, MT., Chao, YC. and Yu, SM. (1994). J. Biol. Chem., **269**, 17635-17641.

Chen, MH., Liu, LF., Chen, YR., Wu, HK. and Yu, SM. (1994). Plant J., **6**, 625-636.

Daussant, J., Zbaszyniak, B., Sadowski, J. and Wiatroszak, I. (1981). Planta, **151**, 176-179.

Denyer, K., Clarke, B., Hylton, C., Tatge, H. and Smith, A.M. (1996). Plant J., **10**, 1135-1143.

Duwenig, E., Steup, M., Willmitzer, L. and Kossmann, J. (1997). Plant J., **12**, 323-333.

Eggermont, K., Goderis, I. J. and Broekaert, W. F. (1996). Plant Mol. Biol. Rep. **14**, 273-279.

Feinberg, A. P. and Vogelstein, B. (1983). Anal. Biochem. **132**, 6-13.

Gelvin, S. B. and Schilperoort, R. A. (1995). Plant Molecular

Biology Manual. 2nd edition. Kluwer Academic Publishers, The Netherlands.

Hildebrand, D.F. and Hymowitz, T. (1981). Physiol. Plant., **53**, 429-434.

Horsch, R.B., Fry, J.E., Hoffman, N.L., Eichholtz, D., Rogers, S.G. and Swaley, R.T. (1985) Science, 227. 1229-1231.

Hylton, C.M., Denyer, K., Keeling, P.L., Chang, MT. and Smith, A.M. (1995). Planta, **198**, 230-237.

Innes, M.A., Gelfand, D.H., Sninsky, J.J. and White, T.J. (1990). PCR Protocols. Publisher: Academic Press.

James, M.G., Robertson, D.S. and Myers, A.M. (1995). Plant Cell, **7**, 417-429.

Jefferson, R.A., Kavanagh, T.A. and Bevan, M.W. (1987) EMBO, J. 6. 3901-3907.

Kakefuda, G., Duke, S.H. and Hostak, M.H. (1986). Planta, **168**, 175-182.

Klosgen, R.B. and Weil, J.H. (1991) Mol. Gen. Genet., 225, 297-304

Li, B., Servaites, J.C. and Geiger, D.R. (1992). Plant Physiol., **98**, 1277-1284.

Liu, X.J., Prat, S., Willmitzer, L. and Frommer, W.B. (1990) Mol. Gen. Genet., 223, 401-406.

Mita, S., Suzuki-Fujii, K. and Nakamura, K. (1995) Plant Physiol. 107,895-904.

Mita, S., Murano, N., Akaike, M. and Nakamura, K. (1997). Plant

J., **11**, 841-851.

Mould, R.M. and Gray, J.C. (1997a). In Cell Biology: A Laboratory Handbook, second edition, Volume 2. (Cells, J.E. ed). New York: Academic Press, pp. 81-86.

Mould, R.M. and Gray, J.C. (1997b). In Cell Biology: A Laboratory Handbook, second edition, Volume 2. (Cells, J.E. ed). New York: Academic Press, pp. 286-292.

Nakamura, K., Ohto, M., Yoshida, N. and Nakamura, K. (1991). Plant Physiol., **96**, 902-909.

Neuhaus, H.E., Henrichs, G. and Schiebe, R. (1995). Planta, **194**, 454-460.

Newman, T. *et al.* (1994). Plant Physiol., **106**, 1241-1255.

Nielson, T.H., Deiting, U. and Stitt, M. (1997). Plant Physiol., **113**, 503-510.

Odell, J.T. Nagy, F. and Chua, N.H. (1985) Nature, 313, 810-812.

Peavey, D.G., Steup, M. and Gibbs, M. (1977). Plant Physiol., **60**, 305-308.

Pwee, K-H. and Gray, J.C. (1993) Plant J. 3, 437-449.

Rocha-Sosa, M., Sonnewald, U., Frommer, W., Stratmann, M., Schell, J. and Willmitzer, L. (1989) EMBO, 8, 23-29.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning. Publisher: Cold Spring Harbour.

Schatz, G. and Dobberstein, B. (1996). Science, **271**, 1519-1526.

Schnell, D.J. and Blobel, G. (1993). J. Cell. Biol., **120**, 103-

115.

Sonnewald, U., Basner, A., Greve, B. and Steup, M. (1995). Plant. Mol. Biol., **27**, 567-576

Sweetlove, L.J., Burrell, M.M. and ap Rees, T. (1996). Biochem. J., **320**, 493-498.

Thomson, W.W. and Whatley, J.M. (1980) Ann. Rev. Plant Physiol. 31, 375-394.

van Engelen, F. A., Molthoff, J. W., Conner, A. J., Nap, J-P., Pereira, A. and Stiekema, W. J. (1995). Transgenic Res. **4**, 288-290.

van der Leij, F.R., Visser, R.G.F., Ponstein, A.S., Jacobsen, E. and Feenstra, W.J. (1991). Mol. Gen. Genet., **228**, 240-248.

Wang, SM., Lue, WL. and Chen, J. (1996). Plant Mol. Biol., **31**, 975-982.

Wang, SM., Lue, WL., Huang, HW. and Chen, J. (1997). Plant Physiol., **113**, 403-409.

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13*bis*)

A. The indications made below relate to the microorganism referred to in the description on page 29, lines 9/10

B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet

Name of depositary institution

The National Collections for Industrial and Marine Bacteria Limited (NCIMB)

Address of depositary institution *(including postal code and country)*

23 St. Machar Drive
Aberdeen AB2 1RY
Scotland, United Kingdom

Date of deposit	Accession Number
4 August 1998	NCIMB 40964

C. ADDITIONAL INDICATIONS *(leave blank if not applicable)* This information is continued on additional sheet

Please find enclosed copies of the Receipt of Deposit and copies of the viability proofs from the Depositary Institution. Name and Address of Depositor: Advanced Technologies (Cambridge) Limited, 210 Cambridge Science Park, Cambridge CB4 0WA, United Kingdom.

- **Characteristics: E. Coli XLI Blue MRF1**
- **Produced by: inserting 1953bp sequence for ct ß-amylase from Arabidopsis thaliana ecotype Columbia inserted into a vector type puC type-pZL1 (Gibco BRL) which is in an E. coli strain XLI Blue MRF1**
- **Usefulness: targeting a particular ß-amylase to a chloroplast**

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE *(if the indications are not for all designated States)*

E. SEPARATE FURNISHING OF INDICATIONS *(leaf blank if not applicable)*

The indications listed below will be submitted to the International Bureau later *(specify the general nature of the indications e.g., "Accession Number of Deposit")*

For receiving Office use only

This sheet was received with the international application ✓

Authorized officer

Clare Lyons

For receiving Office use only

This sheet was received by the international Bureau on:

Authorized officer

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

Advanced Technologies (Cambridge) Ltd.,
210 Cambridge Science Park,
Cambridge.
CB4 4WA

INTERNATIONAL FORM

RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT
issued pursuant to Rule 7.1 by the INTERNATIONAL DEPOSITARY AUTHORITY identified at the bottom of this page

NAME AND ADDRESS OF DEPOSITOR

I. IDENTIFICATION OF THE MICROORGANISM

Identification reference given by the DEPOSITOR:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY:
Escherichia coli (XL1 Blue MRF' pBmy81)	NCIMB 40964

II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION

The microorganism identified under I above was accompanied by:

☐ a scientific description

☒ a proposed taxonomic designation

(Mark with a cross where applicable)

III. RECEIPT AND ACCEPTANCE

This International Depositary Authority accepts the microorganism identified under I above, which was received by it on 4 August 1998 (date of the original deposit)[1]

IV. RECEIPT OF REQUEST FOR CONVERSION

The microorganism identified under I above was received by this International Depositary Authority on (date of the original deposit) and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on (date of receipt of request for conversion)

V. INTERNATIONAL DEPOSITARY AUTHORITY

Name: NCIMB Ltd.,

Address: 23 St Machar Drive,
Aberdeen,
AB24 3RY,
Scotland.

Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorised official(s): Terence Dando

Date: 19 August 1998

[1] Where Rule 6/4(d) applies, such date is the date on which the status of International Depositary Authority was acquired.

Form BP/4 (sole page)

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE

Advanced Technologies (Cambridge) Ltd.,
210 Cambridge Science Park,
Cambridge.
CB4 4WA

NAME AND ADDRESS OF THE PARTY TO WHOM THE VIABILITY STATEMENT IS ISSUED

INTERNATIONAL FORM

VIABILITY STATEMENT
issued pursuant to Rule 10.2 by the
INTERNATIONAL DEPOSITARY AUTHORITY
identified on the following page

I. DEPOSITOR	II. IDENTIFICATION OF THE MICROORGANISM
Name: AS ABOVE Address:	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: NCIMB 40964 Date of the deposit or of the transfer[1]: 4 August 1998
III. VIABILITY STATEMENT	
The viability of the microorganism identified under II above was tested on 8 August 1998 2. On that date, the said microorganism was:	
[3] [X] viable	
[3] [ ] no longer viable	

1 Indicate the date of the original deposit or, where a new deposit or a transfer has been made, the most recent relevant date (date of the new deposit or date of the transfer).

2 In the cases referred to in Rule 10.2(a)(ii) and (iii), refer to the most recent viability test.

3 Mark with a cross the applicable box.

Form BP/9 (first page)

IV. CONDITIONS UNDER WHICH THE VIABILITY TEST HAS BEEN PERFORMED[4]
V. INTERNATIONAL DEPOSITARY AUTHORITY
Name: NCIMB Ltd., Address: 23 St Machar Drive, Aberdeen, A24 3RY, Scotland. Signature(s) of person(s) having the power to represent the International Depositary Authority or of authorised official(s): Terence Dando Date: 19 August 1998

4 Fill in if the information has been requested and if the results of the test were negative.

CLAIMS

1. A nucleic acid sequence known herein as SEQ. ID. No. 1 and being from 1-294 nucleotides and having therewithin a sequence capable of targeting a further coding sequence to a plant plastid, or sequences being at least 65% or more homologous with the disclosed sequence SEQ. ID. No. 1 and having the same targeting ability.
2. A nucleic acid sequence according to Claim 1, wherein said nucleic acid sequence encodes about 94 amino acid residues.
3. A nucleic acid sequence according to Claim 2, wherein said sequence encodes about 85 amino acid residues.
4. A nucleic acid sequence known herein as SEQ. ID. No. 2 and being from 1-1642 nucleotides and having therewithin a sequence capable of encoding ß-amylase, or sequences being at least 65% or more homologous with the disclosed sequence within SEQ. ID. No. 2 and having the same encoding ability.
5. A nucleic acid sequence known herein as SEQ. ID. No. 3 and being from 1-1953 nucleotides and having therewithin a sequence capable of encoding chloroplast targeted ß-amylase, or sequences being at least 65% or more homologous with the disclosed sequence within SEQ. ID. No. 3 and having the same encoding ability.

6. A nucleic acid sequence capable of directing expression of a product encoded by a coding sequence which is operably linked thereto, said nucleic acid sequence being known herein as SEQ. ID. No. 8, or being at least 65% homologous therewith and having substantially the same function thereas, and said nucleic acid sequence being responsive to stimulus, the level of expression of said product being variable in response to the stimulus applied to said nucleic acid sequence.

7. A nucleic acid sequence according to any one of the preceding claims, wherein said nucleic acid sequence is an mRNA, cDNA or a genomic DNA sequence.

8. A method of increasing or decreasing in a plant the activity of an enzyme in the pathway of starch biosynthesis or degradation, the method comprising the steps of stably incorporating into a plant genome a chimaeric gene comprising a nucleic acid sequence encoding a plastid targeting sequence and a coding sequence for an enzyme in the starch biosynthetic or degradative pathway, and regenerating a plant having an altered genome.

9. A method according to Claim 8, wherein said nucleic acid sequence comprises the sequence of Claim 1.

10. A method according to Claim 8, wherein said enzyme in the starch degradative pathway is ß-amylase.

11. A method according to Claim 8, 9 or 10, wherein said

coding sequence is the nucleic acid sequence of Claim 4, being a coding sequence for an enzyme in the starch degradation pathway.

12. A method according to Claim 8, wherein said nucleic acid sequence is the nucleic acid sequence of Claim 5.

13. A method according to any one of Claims 8-12, wherein said enzyme from the starch biosynthetic pathway is one of the group consisting of sucrose synthase, ADPG pyrophosphorylase, starch synthase, branching enzyme, α amylase, isoamylase, non-plastidic ß-amylase, α-glucosidase, starch phosphorylase and disproportionating enzyme.

14. A method of targeting proteins or enzymes to a plant plastid, the method comprising the steps of stably incorporating into a plant genome a chimaeric gene comprising a nucleic acid sequence encoding a plastid targeting sequence and a coding sequence for a protein or an enzyme, and regenerating a plant having an altered genome, the protein or enzyme being one or more in the pathway of the following group: lipid synthesis, photosynthesis, amino acid metabolism, nitrogen fixation, carbon fixation or synthesis of carbohydrate polymers; or being able to confer a characteristic to the plant, the nucleic acid sequence being the nucleic acid sequence of Claim 1.

15. A method according to Claim 14, wherein said characteristics selected from one or more of the following group: herbicide resistance and pest resistance.

16. A method of varying the level of expression of a product encoded by a coding sequence operably linked to a nucleic acid sequence capable of directing expression of said product in a plant, said method comprising the steps of stably incorporating into a plant genome a chimaeric gene comprising a nucleic acid sequence capable of directing expression of a product encoded by a coding sequence that is operably linked thereto, said nucleic acid sequence having substantially the sequence of SEQ. ID. No. 8 or being at least 65% homologous therewith and having substantially the same function thereas, and being responsive to stimulus.

17. A method according to Claim 16, wherein said coding sequence encodes a gene causing sterility in said plant during flowering and which is caused to be expressed by said nucleic acid sequence upon exposure to light.

18. A method according to Claim 16, wherein said coding sequence encodes a gene which alters the level of starch biosynthetic enzyme genes in leaves or other photosynthetic tissue.

19. A method according to Claim 16, wherein said coding sequence encodes a gene whose expression level is varied

upon the generation of sugar in a developing storage organ and the effect of sugar on said nucleic acid sequence.

20. A method according to Claim 16, wherein said nucleic acid sequence is operable when there is light but no sugar.

21. A method according to Claim 16, wherein said nucleic acid sequence is operable when there is sugar but no light.

22. A method of altering in a transgenic plant, which plant already shows an increase or decrease in an enzyme activity as a result of genetic transformation, a further enzyme in order to up or down regulate said further enzyme and thereby increase or decrease the amount of starch produced by the retransformed plant.

23. A method according to Claim 22, wherein said first transformed plant is a plant having an increased or decreased enzyme activity in the starch biosynthetic pathway.

24. A method according to Claim 23, wherein said first transformed plant is a transgenic potato transformed with the gene for ADPG-PPase.

25. A method according to any one of Claims 8, 14 or 16, wherein said chimaeric gene further comprises a promoter selected from the group consisting of the cauliflower mosaic virus 35S promoter (full or truncated), the rubisco promoter, the pea plastocyanin promoter, the nopaline synthase promoter, the chlorophyll r/b binding promoter,

the high molecular weight glutenin promoter, the α, ß-gliadin promoter, the hordein promoter and the patatin promoter.

26. A method according to any one of Claims 8, 14 or 16, wherein said coding sequence of said enzyme in said chimaeric gene provides up or down regulation of the activity of said enzyme.

27. A method according to Claim 16, wherein said stimulus is the presence or absence of light and/or varying levels of sugar.

28. A method according to Claim 16, wherein said stimulus is a stimulus which is developmentally controlled.

29. A method according to Claim 27, wherein the sugar is one or more of sucrose or glucose.

30. A method according to Claim 29, wherein the sugar is sucrose.

31. A chimaeric gene comprising a nucleic acid sequence according to Claim 1.

32. A chimaeric gene comprising a nucleic acid sequence encoding a plastid targeting sequence according to Claim 1 and a nucleic acid coding sequence of an enzyme in the starch, degradative pathway, the enzyme being encoded by the sequence according to Claim 4.

33. A chimaeric gene comprising a nucleic acid sequence capable of directing expression of a further coding

sequence, said coding sequence being the sequence according to Claim 4.

34. A chimaeric gene comprising a nucleic acid sequence according to Claim 6 that is responsive to stimulus and a coding sequence, the level of expression of said coding sequence being variable in response to the stimulus applied to said nucleic acid sequence.

35. Plant cells comprising the nucleic acid sequence of any one of Claims 1, 4, 5 or 6.

36. Seed of the transformed plant containing one or more of the nucleic acid sequences of Claims 1, 4, 5, or 6.

37. Plants transformed according to the methods of any one of Claims 8, 14 or 16, where in said plants are one or more of the group consisting of potato, wheat, maize, barley, tomato, rice, pea, soybean, peanut, cassava, yam, banana and tobacco.

Fig.1.

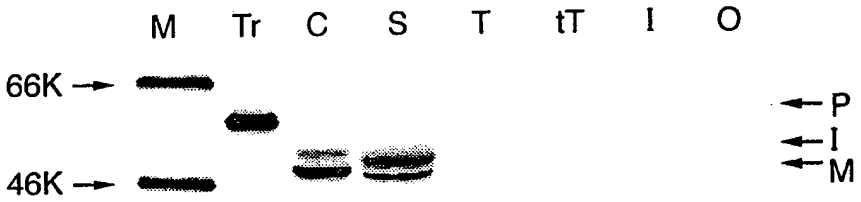


Fig.2(a).

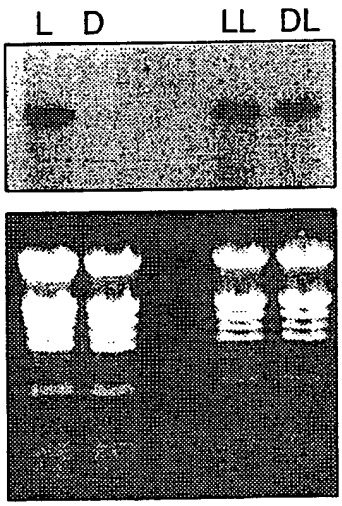
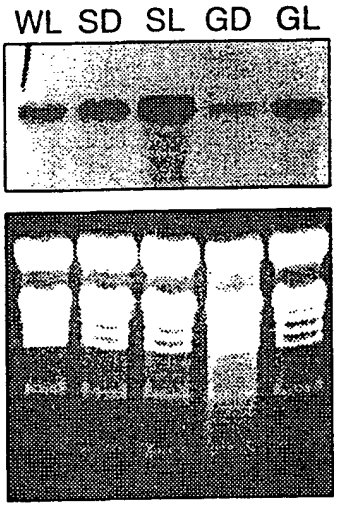


Fig.2(b).



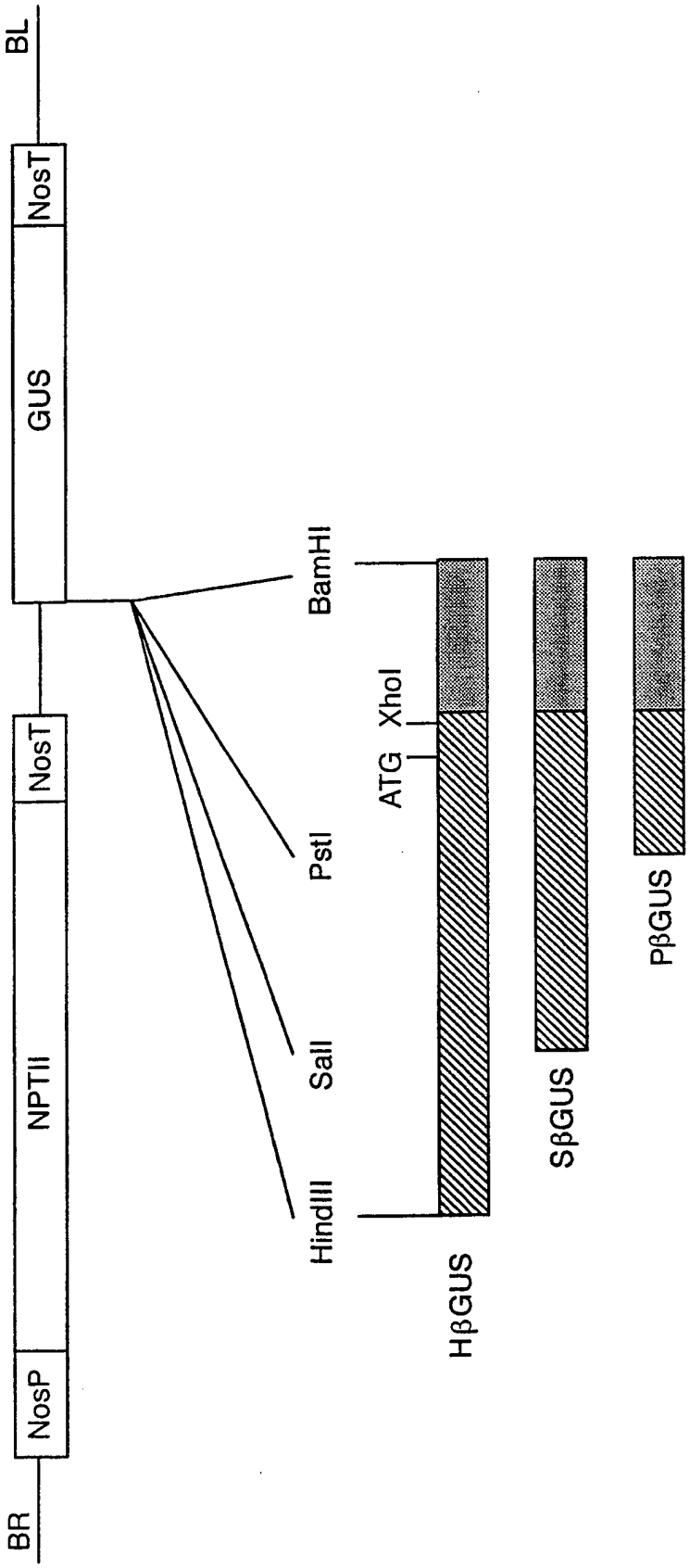

BL
NosT
GUS
NosT
NPTII
NosP
BR
BamHI
PstI
SalI
HindIII
ATG
XhoI
HβGUS
SβGUS
PβGUS

Fig.3.

Fig.4.

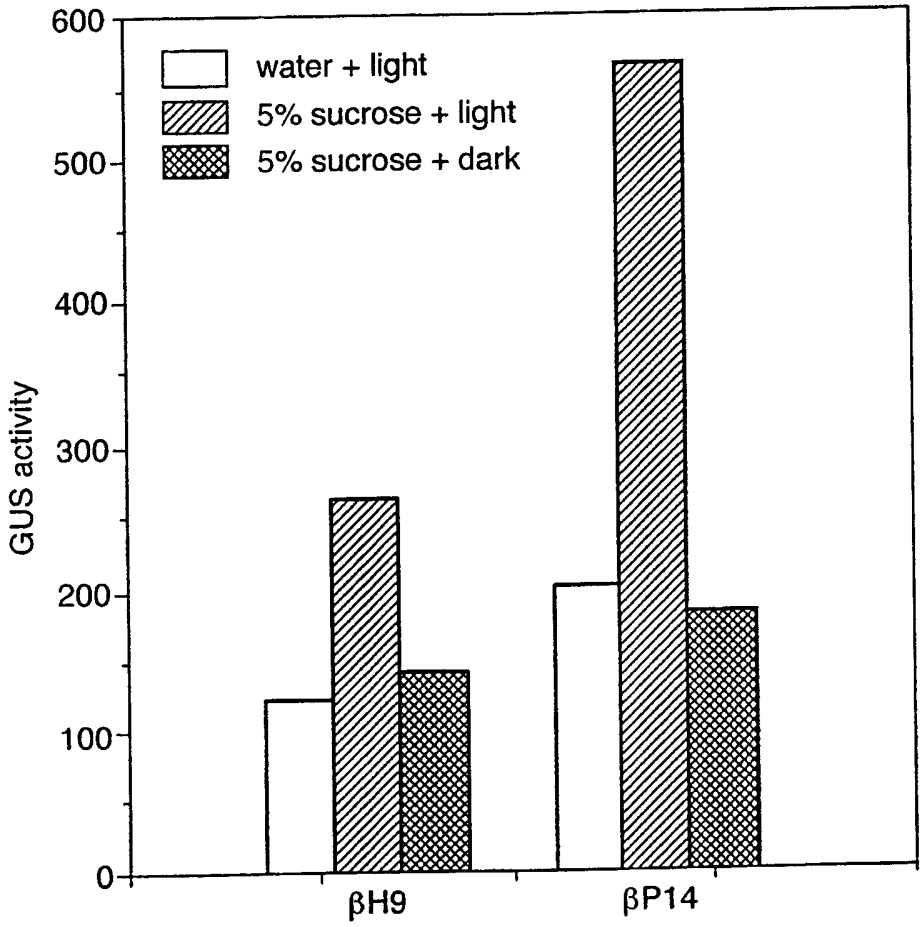

600
500
400
300
200
100
0
GUS activity
water + light
5% sucrose + light
5% sucrose + dark
βH9
βP14

Fig.5.

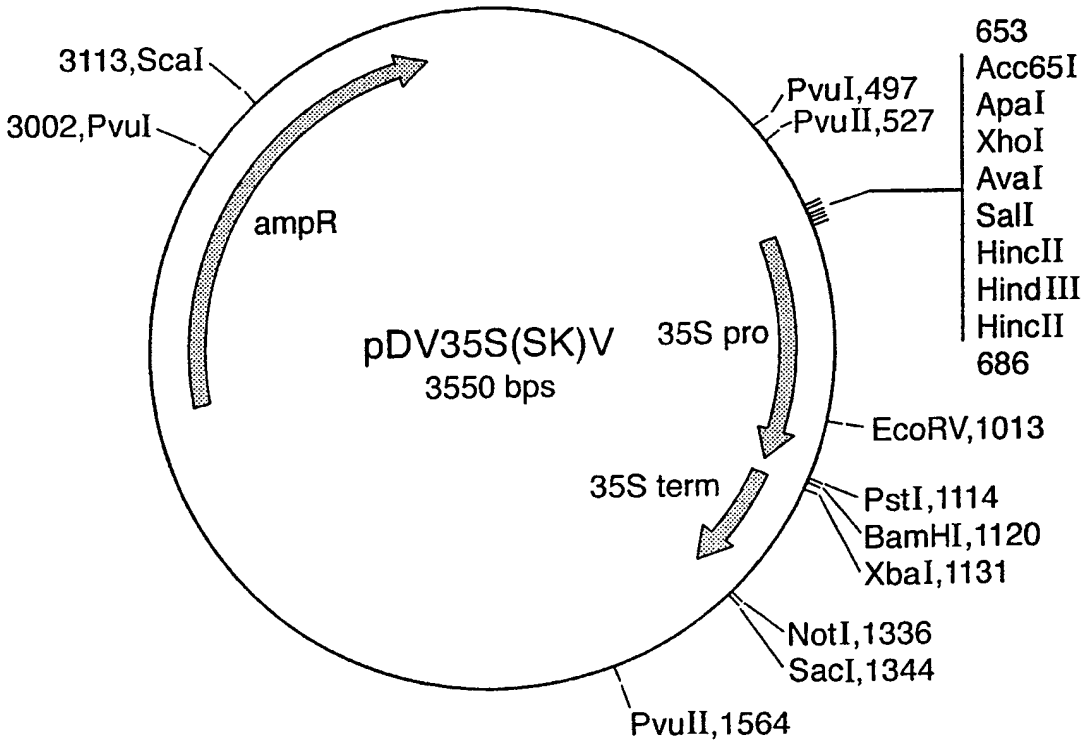

3113,ScaI
3002,PvuI
ampR
pDV35S(SK)V
3550 bps
35S pro
35S term
PvuI,497
PvuII,527
653
Acc65I
ApaI
XhoI
AvaI
SalI
HincII
HindIII
HincII
686
EcoRV,1013
PstI,1114
BamHI,1120
XbaI,1131
NotI,1336
SacI,1344
PvuII,1564

Fig.6.

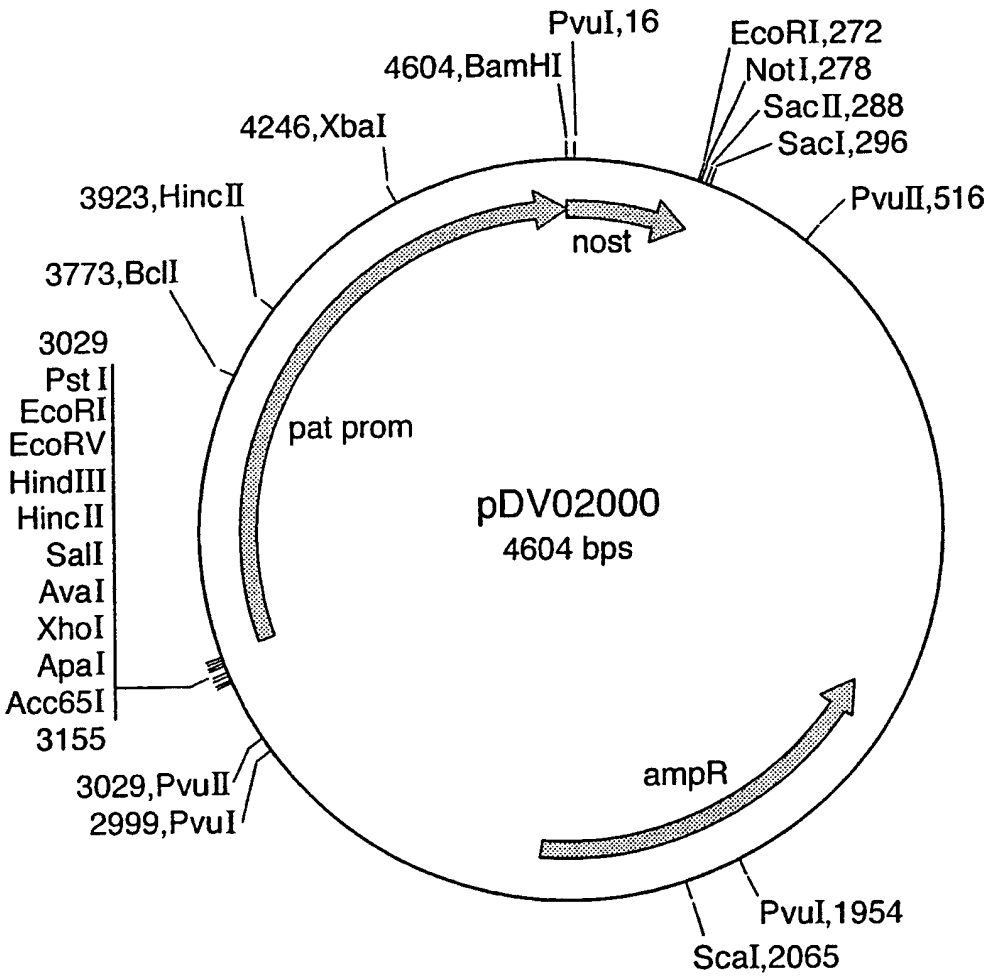

PvuI,16
4604,BamHI
4246,XbaI
3923,HincII
3773,BclI
3029
Pst I
EcoRI
EcoRV
HindIII
HincII
SalI
AvaI
XhoI
ApaI
Acc65I
3155
3029,PvuII
2999,PvuI
EcoRI,272
NotI,278
SacII,288
SacI,296
PvuII,516
nost
pat prom
pDV02000
4604 bps
ampR
PvuI,1954
ScaI,2065

Fig.7.

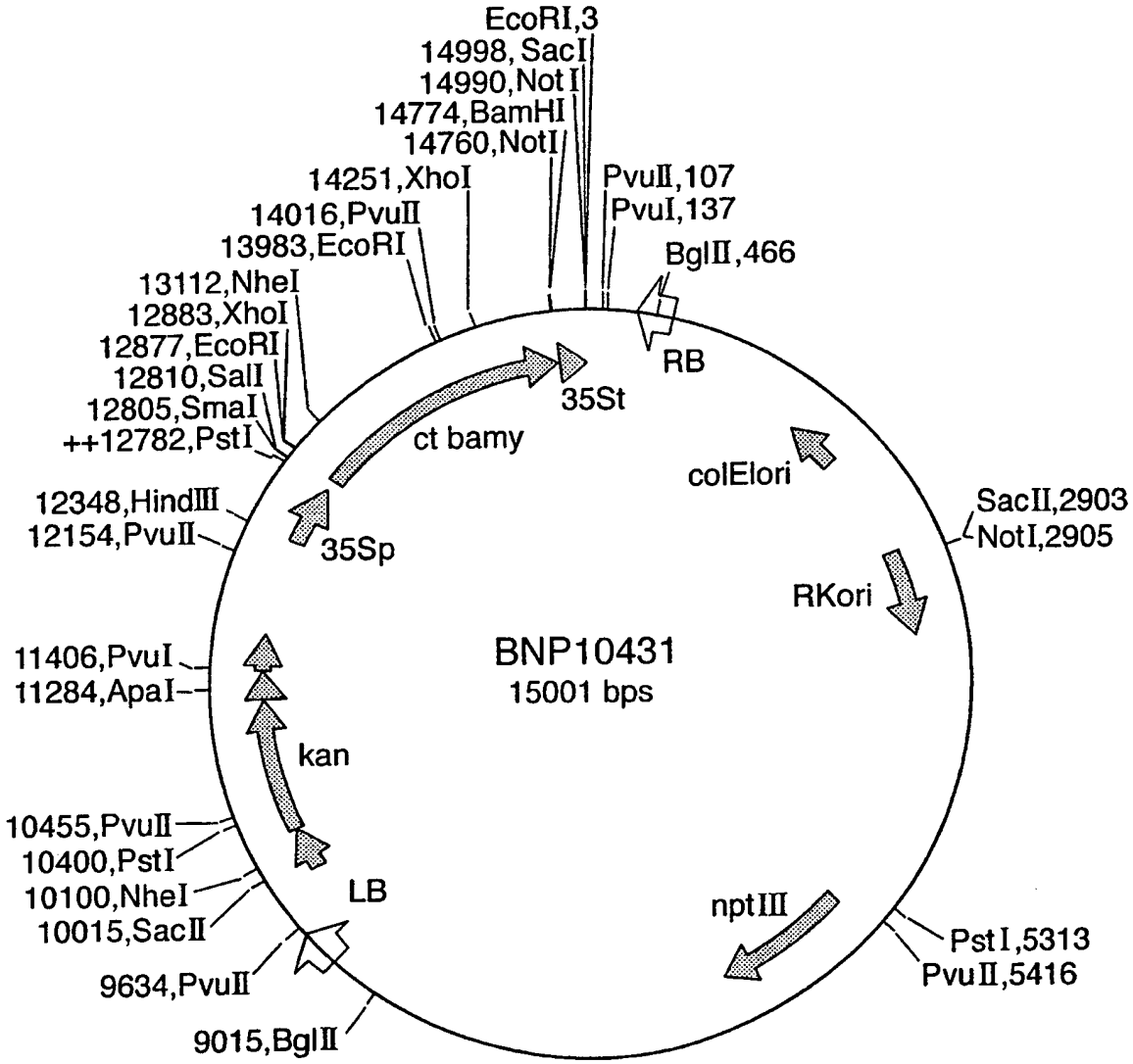

EcoRI,3
14998, SacI
14990,Not I
14774,BamHI
14760,NotI
14251,XhoI
14016,PvuII
13983,EcoRI
13112,NheI
12883,XhoI
12877,EcoRI
12810,SalI
12805,SmaI
++12782,PstI
12348,HindIII
12154,PvuII
11406,PvuI
11284,ApaI
10455,PvuII
10400,PstI
10100,NheI
10015,SacII
9634,PvuII
9015,BglII
PvuII,107
PvuI,137
BglII,466
RB
35St
ct bamy
colElori
SacII,2903
NotI,2905
RKori
BNP10431
15001 bps
35Sp
kan
LB
nptIII
PstI,5313
PvuII,5416

Fig.8.

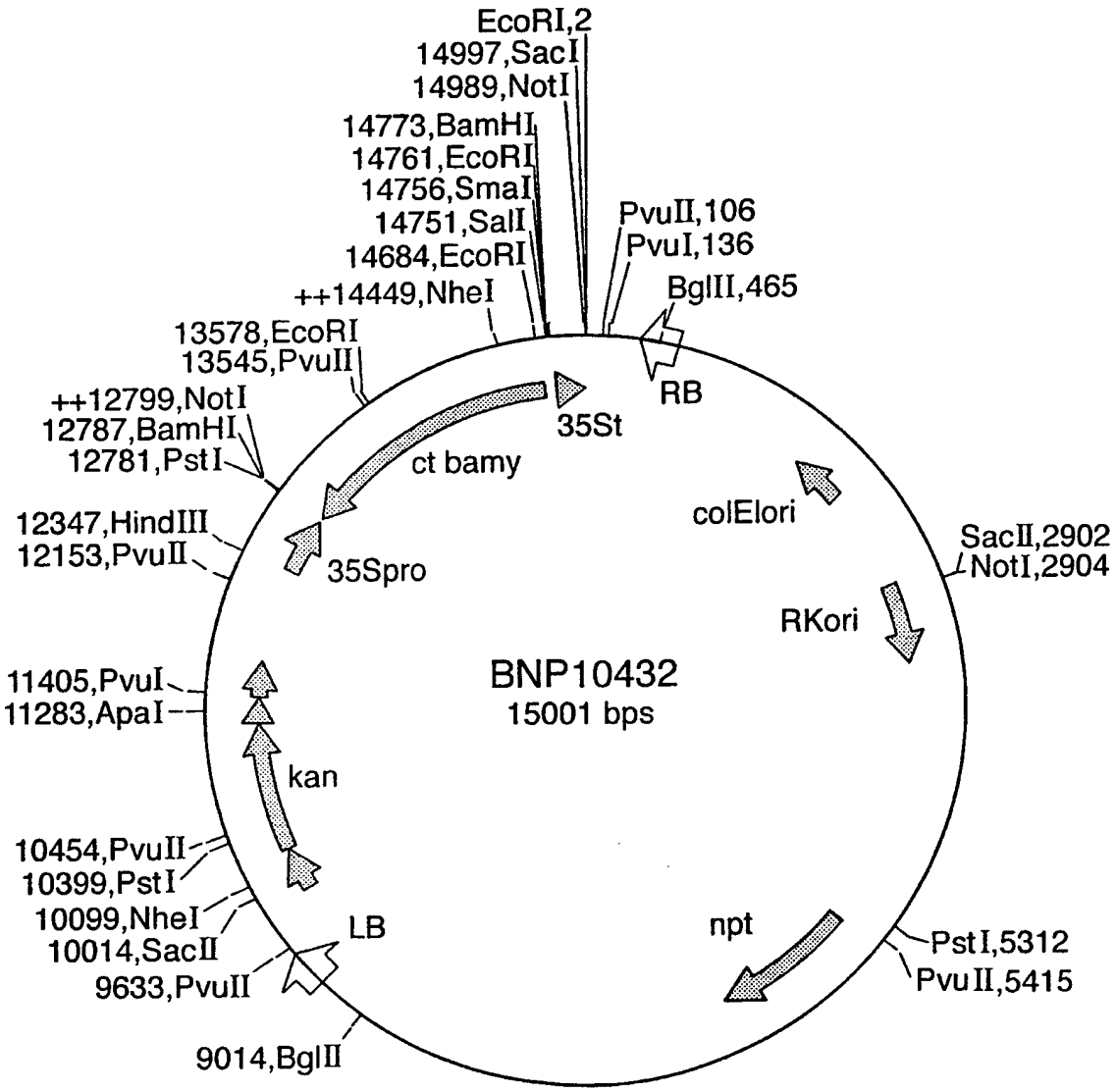

EcoRI,2
14997,SacI
14989,NotI
14773,BamHI
14761,EcoRI
14756,SmaI
14751,SalI
14684,EcoRI
++14449,NheI
13578,EcoRI
13545,PvuII
++12799,NotI
12787,BamHI
12781,PstI
12347,HindIII
12153,PvuII
11405,PvuI
11283,ApaI
10454,PvuII
10399,PstI
10099,NheI
10014,SacII
9633,PvuII
9014,BglII
PvuII,106
PvuI,136
BglII,465
RB
35St
ct bamy
35Spro
colElori
SacII,2902
NotI,2904
RKori
BNP10432
15001 bps
kan
LB
npt
PstI,5312
PvuII,5415

Fig.9.

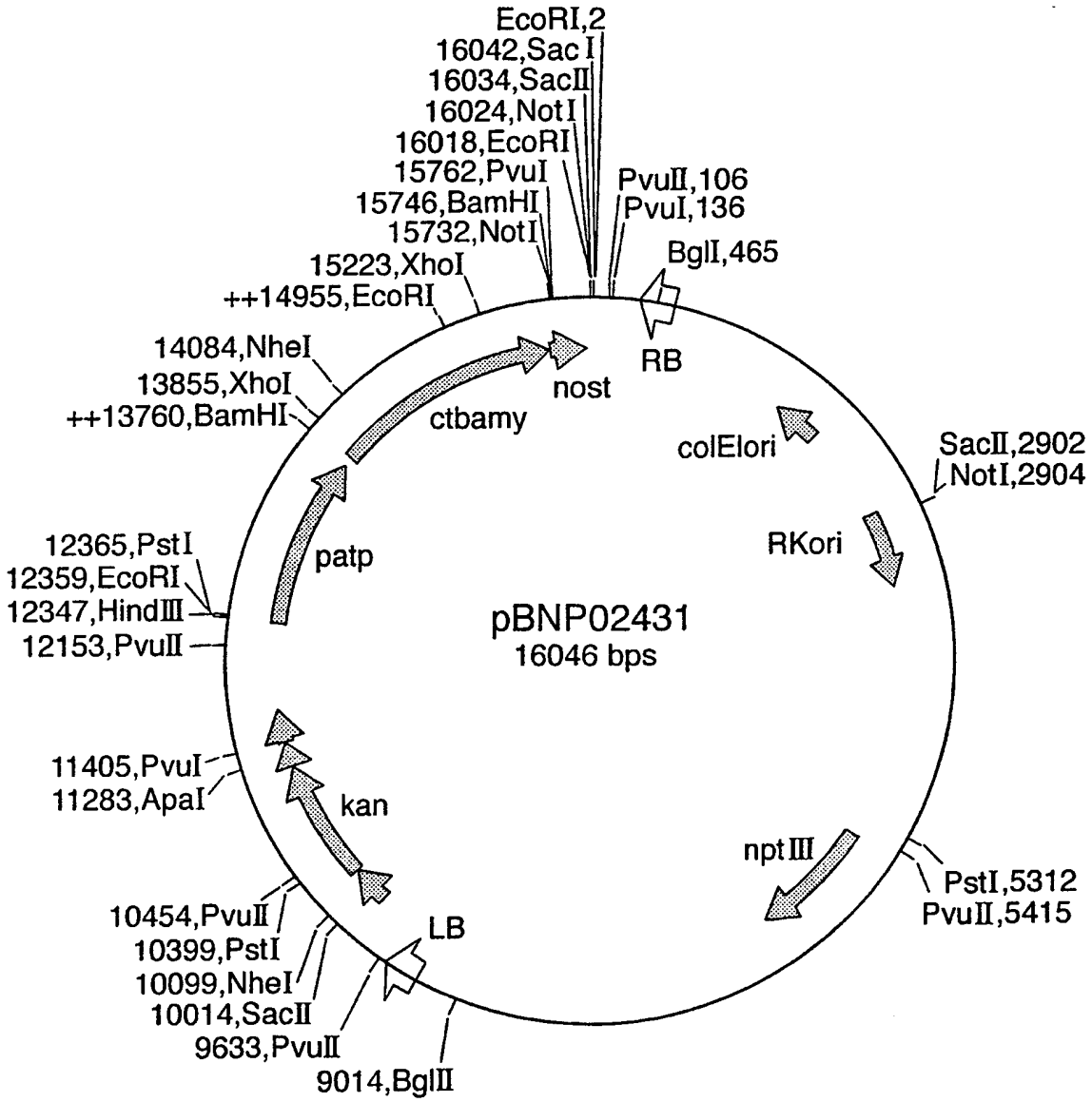
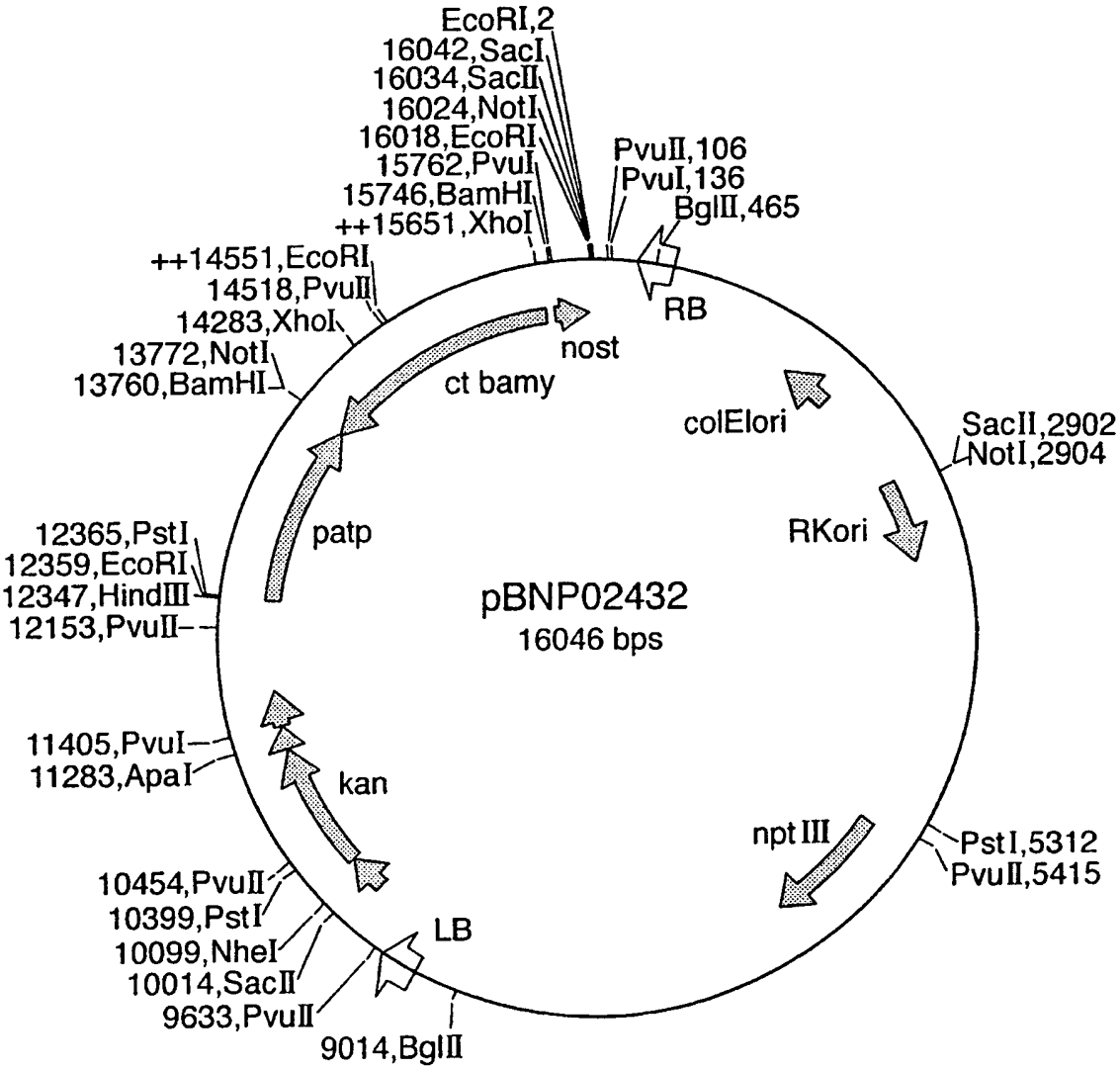

EcoRI,2
16042,Sac I
16034,SacII
16024,Not I
16018,EcoRI
15762,PvuI
15746,BamHI
15732,NotI
15223,XhoI
++14955,EcoRI
14084,NheI
13855,XhoI
++13760,BamHI
12365,PstI
12359,EcoRI
12347,HindIII
12153,PvuII
11405,PvuI
11283,ApaI
10454,PvuII
10399,PstI
10099,NheI
10014,SacII
9633,PvuII
9014,BglII
PvuII,106
PvuI,136
BglI,465
RB
nost
ctbamy
colElori
SacII,2902
NotI,2904
RKori
patp
pBNP02431
16046 bps
kan
nptIII
PstI,5312
PvuII,5415
LB

Fig.10.


EcoRI,2
16042,SacI
16034,SacII
16024,NotI
16018,EcoRI
15762,PvuI
15746,BamHI
++15651,XhoI
PvuII,106
PvuI,136
BglII,465
RB
++14551,EcoRI
14518,PvuII
14283,XhoI
13772,NotI
13760,BamHI
nost
ct bamy
colElori
SacII,2902
NotI,2904
RKori
12365,PstI
12359,EcoRI
12347,HindIII
12153,PvuII
patp
pBNP02432
16046 bps
11405,PvuI
11283,ApaI
kan
nptIII
PstI,5312
PvuII,5415
10454,PvuII
10399,PstI
10099,NheI
10014,SacII
9633,PvuII
LB
9014,BglII

SEQUENCE LISTING

(1) **GENERAL INFORMATION**

(i) **APPLICANTS:**

(A) NAME:	Advanced Technologies (Cambridge) Limited
(B) STREET:	Globe House, 1 Water Street
(C) CITY:	London
(E) COUNTRY:	England
(F) POSTAL CODE:	WC2R 3LA

(ii) **TITLE OF INVENTION:** A novel plastid-targeting nucleic acid sequence, a novel β-amylase sequence, a stimulus-responsive promoter and uses thereof.

(iii) **NUMBER OF SEQUENCES:** 8

(iv) **CORRESPONDENCE ADDRESS:**

(A) ADDRESSEE:	British American Tobacco (Investments) Limited
(B) STREET:	Regents Park Road
(C) CITY:	Southampton
(D) STATE:	Hampshire
(E) COUNTRY:	England
(F) POSTAL CODE:	SO15 8TL

(v) **COMPUTER READABLE FORM:**

(A) MEDIUM TYPE:	Diskette 3.50 inch
(B) COMPUTER:	Viglen P5/75
(C) OPERATING SYSTEM:	MS-DOS Windows 95
(D) SOFTWARE:	Microsoft Word 97

(vi) **CURRENT APPLICATION DATA:**

(A) APPLICATION NUMBER:	Not yet known
(C) CLASSIFICATION:	Not yet known

(viii) **ATTORNEY/AGENT INFORMATION:**

(A) NAME:	Mrs. M.R. Walford/ Mr. K.J.H. MacLean
(C) REFERENCE:	RD-ATC-18

(ix) **TELECOMMUNICATION INFORMATION:**

(A) TELEPHONE:	01703 777155
(B) TELEFAX:	01703 779856

(2) **INFORMATION FOR SEQ. ID. NO:1**

(i) **SEQUENCE CHARACTERISTICS**

(A) LENGTH:	294 bps
(B) TYPE:	Nucleotide
(C) STRANDEDNESS:	Double
(D) TOPOLOGY:	Linear

(ii) **MOLECULE TYPE:** cDNA to mRNA

(vi) **ORIGINAL SOURCE:**

(A) ORGANISM:	Arabidopsis thaliana
(B) STRAIN:	Ecotype Colombia
(F) TISSUE TYPE:	Leaf

(vii) **IMMEDIATE SOURCE:**

(A) LIBRARY:	ABRC DNA Stock Centre
(B) CLONE:	EST 81E10T7

(viii) **POSITION IN GENOME:**

(A) CHROMOSOME/SEGMENT: Chromosome IV

(ix) **FEATURE:**

(A) NAME:	Transit peptide
(B) LOCATION:	37-291 bp

(xi) **SEQUENCE DESCRIPTION:** SEQ. ID. NO:1:

```
1     TCATTTCTCATCATAACAAAGAGAGAGAAAAAAACTATGGAATTGACACTGAATTCCTCG
                                          M  E  L  T  L  N  S  S   8

61    AGTTCTCTTATCAAACGTAAAGATGCCAAGAGTTCTAGAAACCAAGAAAGTTCCTCCAAC
        S  S  L  I  K  R  K  D  A  K  S  S  R  N  Q  E  S  S  S  N  28

121   AACATGACCTTTGCGAAGATGAAGCCGCCAACATATCAGTTCCAAGCAAAGAACTCGGTT
        N  M  T  F  A  K  M  K  P  P  T  Y  Q  F  Q  A  K  N  S  V  48

181   AAGGAAATGAAGTTCACTCACGAGAAGACCTTCACGCCAGAAGGTGAAACCCTTGAGAAA
        K  E  M  K  F  T  H  E  K  T  F  T  P  E  G  E  T  L  E  K  68

241   TGGGAGAAGCTCCACGTTCTCTCATACCCACACTCCAAGAACGACGCTAGCGTT
        W  E  K  L  H  V  L  S  Y  P  H  S  K  N  D  A  S  V        85
```

(2) **INFORMATION FOR SEQ. ID. NO:2**

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH:	1642
(B) TYPE:	Nucleotide
(C) STRANDEDNESS:	Double
(D) TOPOLOGY:	Linear

(ii) **MOLECULE TYPE:** cDNA to mRNA

(vi) **ORIGINAL SOURCE:**

(A) ORGANISM:	Arabidopsis thaliana
(B) STRAIN:	Ecotype Columbia
(C) TISSUE TYPE:	Leaf

(vii) **IMMEDIATE SOURCE:**

(A) LIBRARY:	ABRC DNA Stock Centre
(B) CLONE:	EST 81E10T7

(viii) **POSITION IN GENOME:**

(A) CHROMOSOME/SEGMENT: Chromosome IV

(ix) **FEATURE:**

(A) NAME:	Coding Sequence
(B) LOCATION:	1-1393bps
(D) OTHER INFORMATION:	Mature peptide

(xi) **SEQUENCE DESCRIPTION:** SEQ. ID. NO:2:

```
1     GTTCCGGTGTTCGTCATGTTACCGCTCGACACAGTAACAATGTCAGGGCATTTGAACAAA
        V  P  V  F  V  M  L  P  L  D  T  V  T  M  S  G  H  L  N  K 20

61    CCACGAGCCATGAACGCTAGTTTGATGGCTCTGAAAGGAGCTGGTGTGGAAGGTGTGATG
        P  R  A  M  N  A  S  L  M  A  L  K  G  A  G  V  E  G  V  M 40

121   GTGGATGCTTGGTGGGGATTGGTGGAGAAAGATGGACCTATGAATTATAACTGGGAAGGC
        V  D  A  W  W  G  L  V  E  K  D  G  P  M  N  Y  N  W  E  G 60

181   TATGCCGAGCTTATACAGATGGTTCAAAAGCACGGTCTCAAACTCCAGGTCGTTATGTCA
        Y  A  E  L  I  Q  M  V  Q  K  H  G  L  K  L  Q  V  V  M  S 80

241   TTCCATCAATGTGGAGGAAACGTAGGAGACTCTTGCAGTATCCCCTTGCCTCCATGGGTG
        F  H  Q  C  G  G  N  V  G  D  S  C  S  I  P  L  P  P  W  V 100

301   CTTGAAGAGATCAGCAAGAACCCTGATCTTGTCTACACAGACAAATCTGGGAGAAGGAAC
        L  E  E  I  S  K  N  P  D  L  V  Y  T  D  K  S  G  R  R  N 120

361   CCTGAATATATCTCCTTGGGATGTGATTCTGTGCCTGTCCTAAGAGGAAGAACACCTATC
        P  E  Y  I  S  L  G  C  D  S  V  P  V  L  R  G  R  T  P  I 140

421   CAGGTCTACTCAGATTTCATGAGGAGCTTCCGTGAACGATTTGAAGGCTACATAGGAGGA
        Q  V  Y  S  D  F  M  R  S  F  R  E  R  F  E  G  Y  I  G  G 160
```

```
481  GTTATTGCGGAAATTCAAGTAGGAATGGGACCTTGTGGAGAATTGAGATACCCATCATAC
       V  I  A  E  I  Q  V  G  M  G  P  C  G  E  L  R  Y  P  S  Y 180

541  CCTGAAAGCAACGGGACCTGGAGATTCCCCGGAATTGGAGAGTTCCAGTGCTACGACAAG
       P  E  S  N  G  T  W  R  F  P  G  I  G  E  F  Q  C  Y  D  K 200

601  TATATGAAATCGTCACTTCAGGCATATGCTGAATCAATTGGGAAAACTAACTGGGGAACA
       Y  M  K  S  S  L  Q  A  Y  A  E  S  I  G  K  T  N  W  G  T 220

661  AGCGGACCTCATGATGCCGGCGAGTACAAGAACCTCCCAGAAGATACTGAATTTTTCAGG
       S  G  P  H  D  A  G  E  Y  K  N  L  P  E  D  T  E  F  F  R 240

721  AGAGACGGAACATGGAATAGCGAGTATGGAAAGTTTTTCATGGAATGGTACTCCGGGAAG
       R  D  G  T  W  N  S  E  Y  G  K  F  F  M  E  W  Y  S  G  K 260

781  CTGCTAGAACATGGAGACCAACTCCTATCTTCAGCGAAAGGTATCTTTCAAGGAAGCGGA
       L  L  E  H  G  D  Q  L  L  S  S  A  K  G  I  F  Q  G  S  G 280

841  GCAAAGCTATCAGGAAAGGTAGCTGGAATTCACTGGCACTACAACACCAGGTCACACGCA
       A  K  L  S  G  K  V  A  G  I  H  W  H  Y  N  T  R  S  H  A 300

901  GCTGAGCTAACCGCTGGATATTACAACACAAGAAACCATGACGGGTATCTGCCAATAGCT
       A  E  L  T  A  G  Y  Y  N  T  R  N  H  D  G  Y  L  P  I  A 320

961  AAGATGTTCAACAAACATGGAGTTGTGCTCAACTTCACCTGCATGGAGATGAAAGACGGG
       K  M  F  N  K  H  G  V  V  L  N  F  T  C  M  E  M  K  D  G 360

1021 GAGCAACCTGAGCACGCGAATTGCTCACCAGAAGGTCTGGTCAAGCAAGTACAGAACGCG
       E  Q  P  E  H  A  N  C  S  P  E  G  L  V  K  Q  V  Q  N  A 380

1081 ACAAGGCAGGCCGGAACCGAACTAGCAGGGGAGAACGCGCTAGAACGATATGACTCGAGC
       T  R  Q  A  G  T  E  L  A  G  E  N  A  L  E  R  Y  D  S  S 400

1141 GCATTCGGACAAGTGGTAGCAACAAATAGGTCAGATTCTGGAAATGGGTTAACCGCATTT
       A  F  G  Q  V  V  A  T  N  R  S  D  S  G  N  G  L  T  A  F 420

1201 ACTTACCTAAGAATGAACAAGCGGTTATTTGAGGGTCAAAATTGGCAGCAGTTAGTGGAG
       T  Y  L  R  M  N  K  R  L  F  E  G  Q  N  W  Q  Q  L  V  E 440

1261 TTTGTTAAGAACATGAAGGAAGGTGGTCATGGGAGGAGACTCTCAAAAGAAGACACAACT
       F  V  K  N  M  K  E  G  G  H  G  R  R  L  S  K  E  D  T  T 460

1321 GGAAGTGACCTTTATGTTGGATTTGTCAAAGGCAAGATCGCTGAGAATGTGGAGGAGGCT
       G  S  D  L  Y  V  G  F  V  K  G  K  I  A  E  N  V  E  E  A 480

1381 GCTTTAGTGTAATTTCCCACATAGGTACATACATATAGTGTGGTGTTTATTGTATTCCTG
       A  L  V  -                                                  483

1441 TCTGATAAATAACTAGAGAGATCAAACCAGTAAGAGTGTTAAAGCTATAGATTTGCACAA

1501 TTCTGGGTCAGAGTCAGAGCAAAGAGAAGCAAAATCAAGATGATGTACACTTAGATGTAT

1561 CCTATGAGTTTTCCTTGTACATCATCTTCATACTCTTAATCTCAAATACTATGCATTTTT

1621 CTCCAAAAAAAAAAAAAAAAAGGGCGGCCGCTCTAGAGGATCC
```

(2) INFORMATION FOR SEQ. ID. NO:3

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	1953
(B) TYPE:	Nucleotide
(C) STRANDEDNESS:	Double
(D) TOPOLOGY:	Linear

(ii) MOLECULE TYPE: cDNA to mRNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM:	Arabidopsis thaliana
(B) STRAIN:	Ecotype Columbia
(F) TISSUE TYPE:	Leaf

(vii) IMMEDIATE SOURCE:

(A) LIBRARY:	ABRC DNA Stock Centre
(B) CLONE:	EST 81E10T7

(viii) POSITION IN CHROMOSOME:

(A) CHROMOSOME/SEGMENT:	Chromosome IV

(ix) FEATURE:

(A) NAME:	CDS
(B) LOCATION:	37-1683 bp
(C) OTHER INFORMATION:	Chloroplast targeted

(xi) SEQUENCE DESCRIPTION: SEQ. ID. NO:3:

```
1     TCATTTCTCATCATAACAAAGAGAGAGAAAAAAACTATGGAATTGACACTGAATTCCTCG
                                          M  E  L  T  L  N  S  S   8

61    AGTTCTCTTATCAAACGTAAAGATGCCAAGAGTTCTAGAAACCAAGAAAGTTCCTCCAAC
        S  S  L  I  K  R  K  D  A  K  S  S  R  N  Q  E  S  S  S  N  28

121   AACATGACCTTTGCGAAGATGAAGCCGCCAACATATCAGTTCCAAGCAAAGAACTCGGTT
        N  M  T  F  A  K  M  K  P  P  T  Y  Q  F  Q  A  K  N  S  V  48

181   AAGGAAATGAAGTTCACTCACGAGAAGACCTTCACGCCAGAAGGTGAAACCCTTGAGAAA
        K  E  M  K  F  T  H  E  K  T  F  T  P  E  G  E  T  L  E  K  68

241   TGGGAGAAGCTCCACGTTCTCTCATACCCACACTCCAAGAACGACGCTAGCGTTCCGGTG
        W  E  K  L  H  V  L  S  Y  P  H  S  K  N  D  A  S  V  P  V  88

301   TTCGTCATGTTACCGCTCGACACAGTAACAATGTCAGGGCATTTGAACAAACCACGAGCC
        F  V  M  L  P  L  D  T  V  T  M  S  G  H  L  N  K  P  R  A  108

361   ATGAACGCTAGTTTGATGGCTCTGAAAGGAGCTGGTGTGGAAGGTGTGATGGTGGATGCT
        M  N  A  S  L  M  A  L  K  G  A  G  V  E  G  V  M  V  D  A  128

421   TGGTGGGGATTGGTGGAGAAAGATGGACCTATGAATTATAACTGGGAAGGCTATGCCGAG
        W  W  G  L  V  E  K  D  G  P  M  N  Y  N  W  E  G  Y  A  E  148
```

```
481   CTTATACAGATGGTTCAAAAGCACGGTCTCAAACTCCAGGTCGTTATGTCATTCCATCAA
        L  I  Q  M  V  Q  K  H  G  L  K  L  Q  V  V  M  S  F  H  Q 168

541   TGTGGAGGAAACGTAGGAGACTCTTGCAGTATCCCCTTGCCTCCATGGGTGCTTGAAGAG
        C  G  G  N  V  G  D  S  C  S  I  P  L  P  P  W  V  L  E  E 188

601   ATCAGCAAGAACCCTGATCTTGTCTACACAGACAAATCTGGGAGAAGGAACCCTGAATAT
        I  S  K  N  P  D  L  V  Y  T  D  K  S  G  R  R  N  P  E  Y 208

661   ATCTCCTTGGGATGTGATTCTGTGCCTGTCCTAAGAGGAAGAACACCTATCCAGGTCTAC
        I  S  L  G  C  D  S  V  P  V  L  R  G  R  T  P  I  Q  V  Y 228

721   TCAGATTTCATGAGGAGCTTCCGTGAACGATTTGAAGGCTACATAGGAGGAGTTATTGCG
        S  D  F  M  R  S  F  R  E  R  F  E  G  Y  I  G  G  V  I  A 248

781   GAAATTCAAGTAGGAATGGGACCTTGTGGAGAATTGAGATACCCATCATACCCTGAAAGC
        E  I  Q  V  G  M  G  P  C  G  E  L  R  Y  P  S  Y  P  E  S 268

841   AACGGGACCTGGAGATTCCCCGGAATTGGAGAGTTCCAGTGCTACGACAAGTATATGAAA
        N  G  T  W  R  F  P  G  I  G  E  F  Q  C  Y  D  K  Y  M  K 288

901   TCGTCACTTCAGGCATATGCTGAATCAATTGGGAAAACTAACTGGGGAACAAGCGGACCT
        S  S  L  Q  A  Y  A  E  S  I  G  K  T  N  W  G  T  S  G  P 308

961   CATGATGCCGGCGAGTACAAGAACCTCCCAGAAGATACTGAATTTTTCAGGAGAGACGGA
        H  D  A  G  E  Y  K  N  L  P  E  D  T  E  F  F  R  R  D  G 328

1021  ACATGGAATAGCGAGTATGGAAAGTTTTTCATGGAATGGTACTCCGGGAAGCTGCTAGAA
        T  W  N  S  E  Y  G  K  F  F  M  E  W  Y  S  G  K  L  L  E 348

1081  CATGGAGACCAACTCCTATCTTCAGCGAAAGGTATCTTTCAAGGAAGCGGAGCAAAGCTA
        H  G  D  Q  L  L  S  S  A  K  G  I  F  Q  G  S  G  A  K  L 368

1141  TCAGGAAAGGTAGCTGGAATTCACTGGCACTACAACACCAGGTCACACGCAGCTGAGCTA
        S  G  K  V  A  G  I  H  W  H  Y  N  T  R  S  H  A  A  E  L  388

1201  ACCGCTGGATATTACAACACAAGAAACCATGACGGGTATCTGCCAATAGCTAAGATGTTC
        T  A  G  Y  Y  N  T  R  N  H  D  G  Y  L  P  I  A  K  M  F  408

1261  AACAAACATGGAGTTGTGCTCAACTTCACCTGCATGGAGATGAAAGACGGGGAGCAACCT
        N  K  H  G  V  V  L  N  F  T  C  M  E  M  K  D  G  E  Q  P  428

1321  GAGCACGCGAATTGCTCACCAGAAGGTCTGGTCAAGCAAGTACAGAACGCGACAAGGCAG
        E  H  A  N  C  S  P  E  G  L  V  K  Q  V  Q  N  A  T  R  Q  448

1381  GCCGGAACCGAACTAGCAGGGGAGAACGCGCTAGAACGATATGACTCGAGCGCATTCGGA
        A  G  T  E  L  A  G  E  N  A  L  E  R  Y  D  S  S  A  F  G  468

1441  CAAGTGGTAGCAACAAATAGGTCAGATTCTGGAAATGGGTTAACCGCATTTACTTACCTA
        Q  V  V  A  T  N  R  S  D  S  G  N  G  L  T  A  F  T  Y  L  488

1501  AGAATGAACAAGCGGTTATTTGAGGGTCAAAATTGGCAGCAGTTAGTGGAGTTTGTTAAG
        R  M  N  K  R  L  F  E  G  Q  N  W  Q  Q  L  V  E  F  V  K  508

1561  AACATGAAGGAAGGTGGTCATGGGAGGAGACTCTCAAAAGAAGACACAACTGGAAGTGAC
        N  M  K  E  G  G  H  G  R  R  L  S  K  E  D  T  T  G  S  D  528
```

```
1621  CTTTATGTTGGATTTGTCAAAGGCAAGATCGCTGAGAATGTGGAGGAGGCTGCTTTAGTG
        L  Y  V  G  F  V  K  G  K  I  A  E  N  V  E  E  A  A  L  V  548

1681  TAATTTCCCACATAGGTACATACATATAGTGTGGTGTTTATTGTATTCCTGTCTGATAAA

1741  TAACTAGAGAGATCAAACCAGTAAGAGTGTTAAAGCTATAGATTTGCACAATTCTGGGTC

1801  AGAGTCAGAGCAAAGAGAAGCAAAATCAAGATGATGTACACTTAGATGTATCCTATGAGT

1861  TTTCCTTGTACATCATCTTCATACTCTTAATCTCAAATACTATGCATTTTTCTCCAAAAA

1921  AAAAAAAAAAAGGGCGGCCGCTCTAGAGGATCC
```

(2) INFORMATION FOR SEQ. ID. NO:4

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:	21bp
(B) TYPE	Nucleotide
(C) STRANDEDNESS:	Single stranded
(D) TOPOLOGY:	Linear

(xi) SEQUENCE DESCRIPTION: SEQ. ID. No. 4

AATTCCTCGA GTTCTCTTAT C 21

(2) **INFORMATION FOR SEQ. ID. NO:5**

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH:	21bp
(B) TYPE	Nucleotide
(C) STRANDEDNESS:	Single stranded
(D) TOPOLOGY:	Linear

(xi) **SEQUENCE DESCRIPTION:** SEQ. ID. No. 5

GGGGATCCCT GACATTGTTA C 21

2) **INFORMATION FOR SEQ. ID. NO:6**

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH:	35bp
(B) TYPE	Nucleotide
(C) STRANDEDNESS:	Single stranded
(D) TOPOLOGY:	Linear

(xi) **SEQUENCE DESCRIPTION:** SEQ. ID. No. 6

GCTGGTACTC CTGCAGGATC CGGTCCGGAA TTCCC 35

(2) **INFORMATION FOR SEQ. ID. NO:7**

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH:	35bp
(B) TYPE	Nucleotide
(C) STRANDEDNESS:	Single stranded
(D) TOPOLOGY:	Linear

(xi) **SEQUENCE DESCRIPTION:** SEQ. ID. NO: 7

GGGAATTCCG GACCGGATCC TGCAGGCGTA CCAGC 35

SEQUENCE INFORMATION FOR SEQ. ID. NO. 8

(2) **INFORMATION FOR SEQ. ID. NO:8**

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH:	1652 bp
(B) TYPE	Nucleotide
(C) STRANDEDNESS:	Double stranded
(D) TOPOLOGY:	Linear

(ii) **MOLECULE TYPE:** Genomic DNA

(vi) **ORIGINAL SOURCE:**

(A) ORGANISM:	Arabidopsis thaliana
(B) STRAIN:	Ecotype Columbia
(C) TISSUE:	Leaf

(vii) **IMMEDIATE SOURCE:**

(A) LIBRARY:	EU Arabidopsis Genome Project
(B) CLONE:	Cosmid G16599

(viii) **POSITION IN GENOME:**

(A) CHROMOSOME: Chromosome 4

(xi) **FEATURE:**

(A) NAME:	Promoter
(B) LOCATION:	17534-19185bp of G16599 Complementary strand

(xi) **SEQUENCE DESCRIPTION: SEQ. ID. NO: 8**

```
1    AAGCTTGTGT CTATTTCAAA TTCTTGACCG TAGATGTCAC AACATGCATA
51   TATCATTGAA AACAGAGCAA CACAGGAAAC CAAGCATATG TATCTAGATA
101  TACTTAGCAA GACATAACTA TAGTCTTTGA ATCAACATAG GGATTAATGA
151  TAGAGAATGA GGAAGCTCAA GATTTTATAC TCAGTTTCTT ACAAAACAAA
201  TTTCTCTCTA ACTGCAAAAA CACCAATTAG GATTTGAAGA GCGTACCTGT
251  TTGAGTCAAT GTCCAATGTC GTCCCCCCGC CTTCTACATT TCTTAGCCTG
301  CTGAATAAAA GCACAAGCCA AAATGAGAAG GTGCCAAAGG CGATAAGGAT
351  CAATTTCTAC CATTCAAAAA ACTAATGGTG AGAATTAGAA ACGAGAGAAA
401  ACTACTTGTT GAGGAAATAG CCAAAAGCGC AATCTTCGTC ACCTGAATAA
451  AGACCAAACC GTCACTTTCA ATGAGTCAGC AAGAAAAAGA GAGAGAGAGA
501  GAGAGAGATT CTCTATAACA TTTGAGTCGA CATGGATTCT AATGCATCAA
```

551 AAGTCATCTC CAATAAACAA ACACTTGAAA CTCACATGGC TAATAGAACA

601 AGATCAAAGC CTTAAGTATT AAGCATTACA GACACTACTG GCTAACTTTT

651 GACACATGTT CTTAAGTAAC ATAGTATCAA TATCCGTGAA TCACATCGAA

701 CACACACAAC AAGGGCTTAA TGCATCAAAG TCCTGTTATT TCCATATAAC

751 AACATATTTC ATTTACAAAC AGAATGCAGC ATTCAGGCAG TCCAAATGGA

801 AAGGTTGACA AAAAAATATA ATCTTGTAAC TCTACATATA TGGCAGAATG

851 TAATAACCAG GCAAGAAAAA AACAGAATAA ACAGATCAAT GAGTATGATA

901 TAAAAAAAAG TCACAAAGAA TGTGCCACAG TGAACAAGAG GGCCATGAGA

951 AGAAATTTTC AAAGAAAATA TTAGCATTGT TAGAATTTTT TGGGTCAATG

1001 GATCTGTCAG CTGCTTAGTT GGAAAACACA AATCTTACAG GAAGGAAAGT

1051 CCAAGAAAAA GAAAATAAGC AAAGTTAATA GCCACCACAA GAAATTTCAT

1101 ACAGAAATAA TTAAATCGTT GCACTTATCT TCTTATTCAA ACTAAAATCA

1151 AGAGAACTTA ATAATTTTCA GCCACACGAA CCATGTGTTC AAAGCCAAAG

1201 GTGAGAAGCC AAAATTATCA GCTTATCTCC ATTAACAAGG GAAAAGCAAG

1251 ACTAGATTTA AGAGTTCTCT GTAACTAAAA ACTGCAGGAG TGAGTAAGTA

1301 AATAATTCAC CAACAGGAAA ACAAAACTCA ATTATCTATA GCTGAATACA

1351 CATGTAAATG AGAATTTATT AACTAAAACA TCTTCCTTTG TAACTGATGT

1401 GACATTTACA ATTTTTCATT TTGAGGTGTA AGAACCGTGT GACAAGTGAA

1451 AAGGTTAAAA TAAGCAACCT TTGTGATATT TTCTCTCCAC TTTTTGAAAT

1501 TGGGTCTCCA AACCACAGCC AATCAATATT CTTTATAAAT ACAAACACAC

1551 AAACAGCATC TTTCTCTCAA ACACAAACAT ATCTTCTATC AAACACCAAC

1601 AGCTCTATTC TCTACCTCAT TTCTCATCAT AACAAAGAGA GAGAAAAAAA

1651 CT